TESTING AVIATION CADETS FOR PSYCHOMOTOR COORDINATION

UTILIZING HUMAN TALENT

Armed Services Selection and Classification Procedures

BY *Frederick B. Davis*
FOR THE COMMISSION ON IMPLICATIONS OF ARMED SERVICES EDUCATIONAL PROGRAMS

★

AMERICAN COUNCIL ON EDUCATION
Washington, D. C.

PRINTED IN THE UNITED STATES OF AMERICA

FOREWORD

THE COMMISSION on Implications of Armed Services Educational Programs is charged with identifying features of the wartime training and educational enterprises of the Army and Navy which may contribute to the advancement of American education in time of peace.

What the armed services did in the task of classifying personnel and finding the right man for the right job constitutes one area of the investigation, with implications for aptitude testing, guidance and counseling, and for admission and selection policies and practices in schools and colleges. In one sense the function of education may be aptly stated as that of finding and developing human talents, looking toward their optimum utilization in the public interest.

For this particular phase of its studies the Commission engaged the services of Dr. Frederick B. Davis, until recently a major in the Army Air Forces. Dr. Davis was awarded the Legion of Merit for his work in the AAF Aviation Psychology Program developing the AAF Qualifying Examination and other tests used for selecting and classifying aircrew members. Prior to his military service, he was a member of the staff of the Cooperative Test Service of the American Council on Education.

The Secretary of War and the Secretary of the Navy agreed to cooperate in the entire project of the Commission and facilitated its progress by designating as official liaison agencies respectively the Historical Division, War Department Special Staff, and the Standards and Curriculum Division, Training Activity, Bureau of Naval Personnel. These agencies provided full access to documentary materials and entree to numerous armed services headquarters and training installations.

The same agencies also reviewed the studies in manuscript, on occasion gave valuable suggestions, and finally approved the drafts for factual accuracy and to safeguard information vital to the national security. Opinions and assertions contained in the studies are private ones of the author and are not to be

construed as official or as reflecting the views of the War Department or the Navy Department or of the military or naval services at large.

This report touches new frontiers in the development and use of aptitude-testing instruments and procedures, with a constant view to their applicability in the American educational system as it is today and will be in ensuing years. I commend it to students, educators, employers, and all who are interested in more effective utilization of our human and educational resources.

ALONZO G. GRACE
Director

PREFACE

FOR THE convenience of readers of diverse backgrounds and purposes, this report has been separated into three divisions. In chapter i, a brief description of the procedures used to select and classify men and women in the armed forces is presented. For those who have no acquaintance with the mechanics of these procedures, chapter i may serve as enough of an introduction to permit a meaningful study of chapter ii. The major emphasis in this report is on the civilian implications listed and discussed in chapter ii. For clarity in presentation these implications are grouped conveniently under broad implications which have been set as marginal headings in italics. An effort has been made to present in nontechnical terminology implications of practical significance.

Recognizing that the findings would in large part probably not be novel, the members of the Commission on Implications of Armed Services Educational Programs charged the writer with preparing a report that would provide a basis for practical action by school administrators, leaders in guidance, and test constructors. Consequently, the implications are set forth in a form that is frankly and even aggressively evaluative and hortatory. Each of the implications is derived from data accumulated by one or more of the armed services, but this does not exclude the possibility that other interpretations of the same data may legitimately be made.

In two brief appendixes some information of interest mainly to technicians is presented. Appendix A includes a few implications, and Appendix B consists of a discussion of some considerations in the selection of test items or of subtests for an examination used for prediction purposes.

The writer wishes to express his appreciation of the helpful suggestions and comments made by Guy L. Bond, University of Minnesota; Herbert S. Conrad, College Entrance Examination Board; Donald W. Fiske, University of Michigan; Felix Kampschroer, War Department; Truman L. Kelley, Harvard Uni-

versity; Helen R. Haggerty, Navy Department; William G. Mollenkopf, Princeton University; Capt. Boyd C. Shafer, War Department; Dewey B. Stuit, State University of Iowa; and Robert L. Thorndike, Columbia University.

Finally, the encouragement and guidance of Alonzo G. Grace and M. M. Chambers during the course of the study and the helpful suggestions of all of the members of the Commission on Implications of Armed Services Educational Programs were greatly valued. The burden of preparing the manuscript was immeasurably lightened by Sadie Kesselman.

FREDERICK B. DAVIS

Washington, D. C.
September 1, 1946

CONTENTS

I. SELECTION AND CLASSIFICATION PROCEDURES

ARMY PROCEDURES

DURING World War I, the Army developed a personnel-classification system that was commonly regarded as the most satisfactory employed up to that time in military history. This system was developed under the guidance of the Committee on Classification of Personnel in the Army, a group of specialists appointed by the Secretary of War. Its objectives were "to secure a contented and efficient Army by placing each enlisted man where he has the opportunity to make the most of his talent and skill; to commission, assign, and promote officers on merit; and to simplify the procedure of discovering talent and assigning it where most needed." [1]

To attain these objectives, classification tests, trade tests, and qualification cards were designed, constructed, and put into effective use. Job analyses were also made. The most famous of the instruments used in the Army during World War I were the initial general-ability tests, the Army Alpha Examination, and the Army Beta Examination. The data obtained from the use of these tests have provided the basis for hundreds of research studies of various types. The important findings are presented in the *Memoirs of the National Academy of Sciences*.[2]

After World War I, the importance of the proper and continued classification of manpower was, for many reasons, not emphasized in the United States Army. Consequently, little systematic effort was made to determine the special qualifications of enlisted men or officers, or to place them in duties for which they were peculiarly fitted. In general, the guiding principle in making assignments was to follow the individual's own pref-

[1] U. S. War Department, *The Personnel System of the United States Army* (Washington: Government Printing Office, 1919), Foreword.

[2] R. M. Yerkes, "Psychological Examining in the United States Army," *Memoirs of the National Academy of Sciences*, Vol. XV (Washington: Government Printing Office, 1921).

erences. "The administration of classification tests, although prescribed by regulations, became a dead letter." [3]

Although the need for up-to-date selection and classification procedures and instruments had long been recognized, the outbreak of war in Europe in 1939 found the Army without them and without trained personnel technicians able to prepare them. When the President declared a state of limited national emergency on September 8, 1939, and the Congress authorized significant increases in the strength of the Regular Army and the National Guard, it became evident to some far-sighted officers that preparations for selecting and classifying a large number of men should be initiated. As the nature of the war in Poland and in western Europe became clear, realization of the level of technical proficiency required among the officers and men of a modern army emphasized the critical need for proper classification of manpower. The pattern of warfare set by the German armies indicated that technicians in unprecedented numbers would be required to maintain and operate the highly mechanized weapons of a modern army. It became evident that if the proper classification of men had been desirable in World War I, it would be crucial in World War II.

Lacking trained personnel to design and construct the tests required for an adequate personnel-classification system, the War Department in late 1939 and early 1940 called on psychologists for aid in preparing a new classification test that would measure general ability to learn. Plans for the new Army General Classification Test were drawn up and presented in May 1940 to the Committee on Classification of Military Personnel. At this same meeting of the committee a report was made with respect to the first of an Army-wide series of job analyses being done for the War Department by the United States Employment Service.

In July 1940 a new Soldier's Qualification Card (WD AGO Form 20) was released. This forms the basic record which follows the soldier throughout his Army career and on which is re-

[3] H. C. Holdridge, "The Army Personnel System," *Adjutant General's School Lecture Series,* No. 1 (Fort Washington, Md.: The Book Service, The Adjutant General's School, 1942), p. 3.

corded the information that forms the basis for his classification and assignment. Thus, by July 1940, work had been started on three requirements for effective classification: job analyses, aptitude tests, and a cumulative record card.

The magnitude of the task of selection and classification, for which the foundations were laid early in 1940, is indicated by the fact that between June 1939 and June 1945 over 9,750,000 men, exclusive of the National Guard, entered the Army. All of these men, including more than 643,000 officers, plus thousands of National Guard and Regular Army personnel, were individually classified and assigned to duty.

Although the major emphasis of this report is placed on the use of tests and interviews for purposes of selecting and classifying personnel, the primary importance of medical, including psychiatric, examinations in determining assignments should be fully recognized.

ENLISTED MEN

By far the largest number of men came into the Army through the operation of Selective Service. When these men reported at an induction station, their eligibility for induction was determined and they were assigned either to the Army or to the Navy on the basis of their qualifications, their preferences, and the needs of the services at the time. Eligibility for induction was determined by means of interviews and tests. If judged fit for induction by medical and psychiatric examiners, high school graduates were accepted without further examination. Other men were required to take the Qualification Test, which included comprehension of numbers, elementary arithmetic, and reading comprehension. Men who obtained an acceptable score on the Qualification Test were inducted. Those who failed the Qualification Test were sifted still further in order to discover those men who could be expected to become useful members of the Army after a brief period of elementary school work in a special-training unit. The first test given for this purpose was the Group Target Test, a nonverbal examination intended to measure learning ability. Men who reached the minimum acceptable score on this test were inducted and assigned

to a special-training unit. Those who failed the Group Target Test were given an individual examination to make sure that no man with adequate learning ability for military duty would be rejected because of failure to understand the group-test directions or because of mental confusion resulting from being tested in unfamiliar surroundings. Only men who failed the individual examination were not inducted.

From the induction station, all men were sent to a reception center. Here, the Army General Classification Test, a test of general learning ability, was administered to all literate men. The General Mechanical-Aptitude Test and the Army Radio Code-Aptitude Test were administered when men with certain special qualifications were required for training. Each man was interviewed by trained personnel and the Soldier's Qualification Card (WD AGO Form 20) was filled out. On this card were recorded data regarding the man's education, his main and secondary occupations, and his job history. Standardized Oral Trade Questions were used when necessary to check on the accuracy of the man's statements about his job history. In addition to information regarding his educational and vocational training and experience, the man's hobbies, ability in sports, and talents for entertaining were recorded. Any evidence of leadership was noted. This qualification card formed the cumulative record that was transferred with the man wherever he went in the Army and that formed the basis for his initial assignment to duty and for any subsequent new classification. From this card, a personnel officer could at once obtain important information regarding the man's previous experience, training, and his aptitudes, as shown by test scores. This information together with that obtained from further interviews and conferences, was used for such purposes as locating specialists, assigning personnel to special-training units, and filling requisitions for assignments to field units and replacement training centers. Moreover, the data on the Soldier's Qualification Card were used time and again throughout a man's career in the Army to aid in determining his qualifications for new assignments, including those for reconditioning training after wounds in combat. Most men received preliminary training and further classification

in replacement training centers or field units. Here, a very large number and variety of specialized tests were employed as the need arose. Minimum qualifying scores on these tests were established for admission to certain courses in which they were predictive of success. In AAF replacement training centers, for example, a Surface-Development Test and a Mechanical-Information Test were used to qualify men for courses in aircraft welding, parachute rigging, and photography. Other tests used in replacement training centers included:

Apprentice-Mechanics Test
Auto-Mechanic-Experience Check List
Carpenter-Experience Check List
Carpenter Test
Clerical-Achievement Test
Clerical-Experience Check List
Cooking-Experience Check List
Cooking Test
Dictation Test
Machinist-Experience Check List
Machinist Test
Supply-Clerk-Experience Check List
Supply-Clerk Test
Truck-Driver-Experience Check List
Truck-Driver Test
Typing Test
Welding-Experience Check List
Welding Test

Tests commonly employed in basic training units in the Army Air Forces (which corresponded to replacement training centers in the Army Ground and Service Forces) included:

Clerical-Work Test
Mechanical-Information Test
Mechanical-Movements Test
Radio and Link-Trainer Test
Radio Code Test
Shop-Mathematics Test
Surface-Development Test
Weather Test

When these tests were originally constructed, objective data were not available regarding their correlation with degree of

success in the courses of training for which they were designed to be predictive. Ordinarily, the selection of content of a test was based on an analysis of the curriculum of the course or of the nature of the job for which the test was intended to select personnel. To obtain objective evidence of the actual correlation between degree of success and scores on the tests, the marks obtained by a group of men at the end of the training course or the supervisor's ratings of their competence on the job were obtained and correlated with the scores obtained by the same men when they were tested. The resulting correlation coefficient provided an index of the prediction efficiency of the test for selecting men for a particular course of training or duty assignment. This process of checking on the usefulness of tests is exceedingly important since it forms the basis for weighting the importance of test scores for various purposes and leads to modifications that can result in marked increases in the efficiency of selection and classification procedures.

OFFICERS

By far the largest proportion of men commissioned in the Army of the United States during World War II were graduates of officer-candidate schools. Men were selected for entrance to these schools from among applicants who attained scores of at least 110 on the General Classification Test, who were recommended by their commanding officers, and who were approved by boards of officers convened for the purpose by the commanding generals of the service commands. To aid the board in its decision regarding each man, it was permitted to judge each man's leadership ability, to consult his qualification card, and to obtain ratings on various traits of leadership by the man's immediate commanding officer.

The selection of officers for direct commissioning from civilian life was accomplished mainly by the men themselves in applying for the more highly specialized positions. In some cases, civilian specialists were actively recruited and offered commissions.

The classification of officers was done on the basis of their previous training and experience by the same means described for enlisted men. Each man was interviewed individually and

the appropriate data were recorded on a qualification card.

The selection and classification of officers for aircrews in the Army Air Forces constituted a special problem of great complexity which was handled under the direction of the Office of the Air Surgeon. Pilots, bombardiers, and navigators were trained as aviation cadets and were recruited directly from civilian life or from personnel already in the Army. From December 7, 1941, to September 1, 1945, 59,053 aviation cadets received appointments as reserve officers in the Air Corps Reserve and 198,037 received appointments as officers in the Army of the United States. Thus, slightly over one-third of all male officers appointed during the war came from the ranks of aviation cadets.

The selection of aviation cadets began with the administration of the Aviation Cadet Qualifying Examination[4] and a physical examination by Aviation Cadet Examining Boards scattered throughout the continental United States and Army installations overseas. The qualifying examination was an essentially unspeeded paper-and-pencil test designed primarily to determine whether the applicant was of officer caliber and, specifically, to assess his chances of completing pilot training. Since it was possible for an applicant to take the examination every thirty days, new forms had to be made available periodically. This provided an unparalleled opportunity for improving the examination systematically on the basis of research findings regarding its prediction efficiency for selecting men capable of succeeding in pilot training. This was done, and the total scores derived from later forms of the examination administered in Aviation Cadet Examining Boards displayed a positive correlation coefficient slightly over .50 with the criterion of graduation or elimination many months later from advanced pilot training.[5] The extent to which graduation from advanced pilot training was associated with nine different levels of score on Test AC121 of the qualifying examination is illustrated in Figure 1.

[4] The title of this examination was changed in 1944 to AAF Qualifying Examination. See F. B. Davis, ed., *The AAF Qualifying Examination*, AAF Aviation Psychology Program Research Reports, No. 6 (Washington: Government Printing Office, 1947).

[5] These data are based on an essentially unselected sample of applicants for pilot training. No corrections for range or for attenuation have been made.

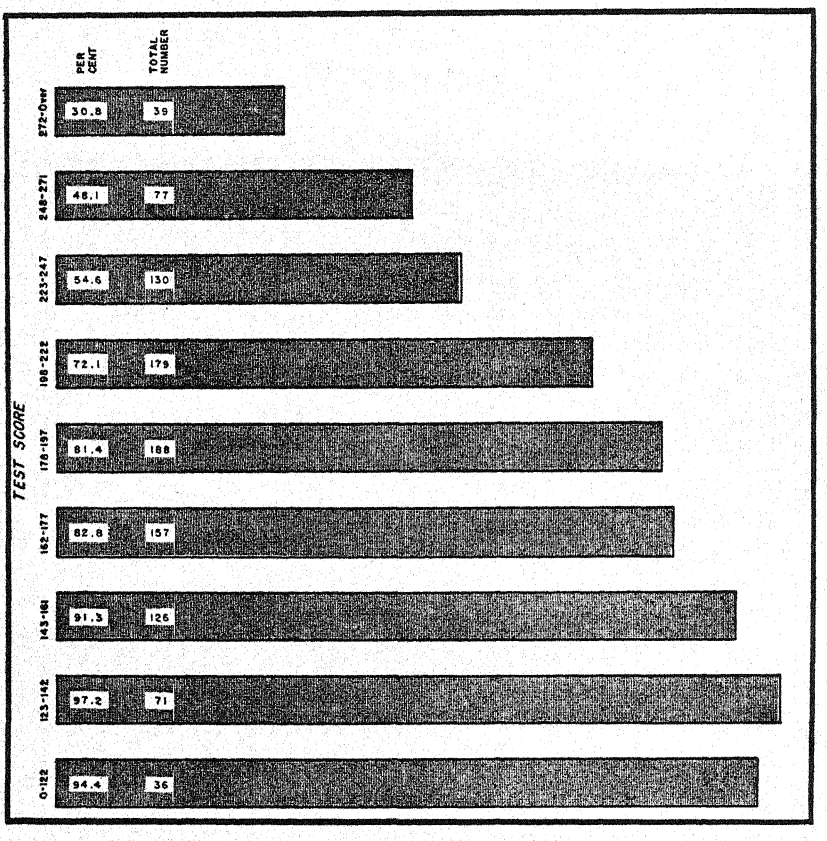


FIG. 1.—Percentage of essentially unselected applicants for pilot training eliminated from preflight through advanced pilot training at each of nine raw-score levels on the Aviation Cadet Qualifying Examination. Based on 1,003 aviation cadets in the AAF, of whom 750 were eliminated. See P. H. DuBois, ed., *The Classification Program,* A.A.F. Aviation Psychology Program Research Reports, No. 2 (Washington: Government Printing Office, 1947), Fig. 5.45.

Following admission to aviation-cadet training and immediately prior to assignment in preflight school, each man was tested for classification purposes with a more elaborate battery of approximately twenty examinations, including psychomotor-apparatus tests. This battery of tests was originally intended to serve primarily as a means of differentiating between men whose aptitudes best fitted them for assignment as pilots, bombardiers, or navigators. As time went on, however, the supply of aircrew officers became greater than was needed, and scores derived from

the Aircrew Classification Battery came to be used merely as a second and more comprehensive selection device. How satisfactory an instrument it was for selecting men who would be likely to graduate from advanced pilot training is shown by the

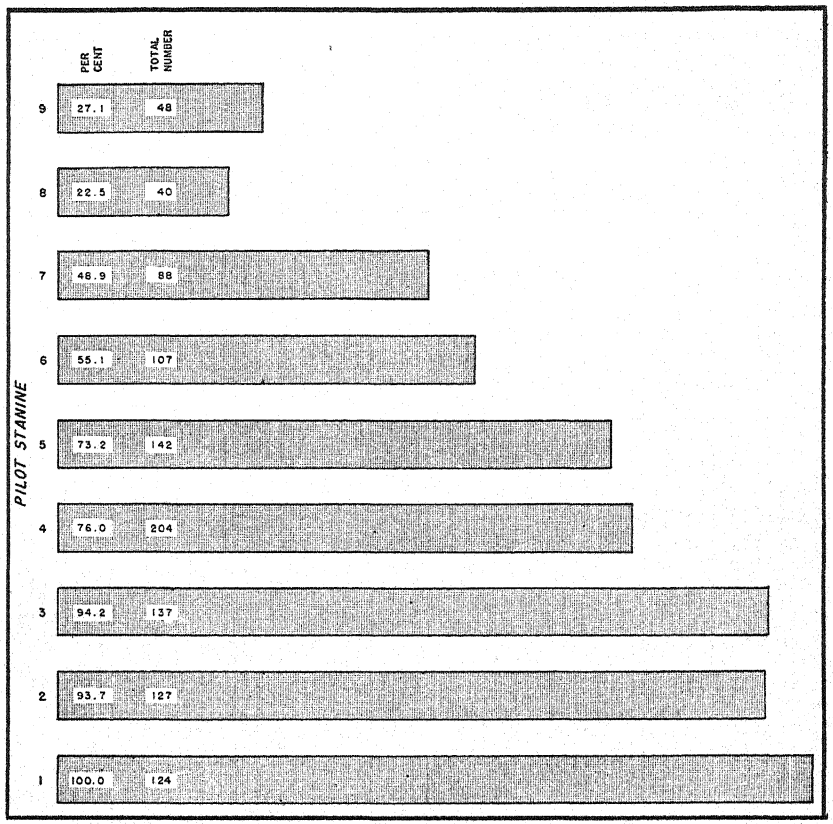


FIG. 2.—Percentage of essentially unselected applicants for pilot training eliminated from preflight through advanced pilot training at each of nine pilot stanine levels. Based on 1,017 aviation cadets in the AAF, of whom 755 were eliminated. Credit for previous flying experience is not included in the stanine scores. See DuBois, *op. cit.*, Fig. 5.43.

data in Figure 2. The biserial coefficient of correlation based on these data is .66. The prediction efficiency of the composite score used to predict graduation or elimination from navigator training was at least as great and perhaps even greater than that of the composite score used to predict graduation or elimi-

nation from pilot training, while the composite score used to predict graduation or elimination from bombardier training was not so effective.

The composite scores derived from the Aircrew Classification Battery were expressed as comparable standard measures known as "stanine scores." Classification of each aviation cadet as a pilot, a navigator, or a bombardier was recommended on the basis of a comparison of his relative standing on these three stanine scores, his preferences, the results of a personal interview, medical data, and the needs of the Air Forces at the time. Since each of the three stanine scores was a composite so weighted as to maximize its correlation with graduation or elimination from one of three types of flying training, the correlation of the stanine scores was dependent in part on the actual overlapping of abilities required for success in each of the three types of flying training. Indications are that the least overlapping of abilities occurs for pilot and navigator functions. In the construction of measuring instruments to be used for purposes of differential classification, it is of considerable importance that systematic efforts be made to measure separately the psychological elements that are unique in each job. This is especially true when selection on the basis of elements likely to be present in all of the jobs has already been made, as in the case of aviation cadets previously selected by means of the qualifying examination.

Warrant officers were selected by means of an examination written especially for the purpose in 1941. Both a general section and a highly technical section were administered to each man. These examinations were prepared under the direction of the Personnel Research Section of the Adjutant General's Office.

WOMEN'S ARMY AUXILIARY CORPS AND WOMEN'S ARMY CORPS PERSONNEL

The selection and classification of personnel for the WAAC and the WAC were carried on in much the same manner as for male officers and enlisted men. However, the selection of the initial group of WAAC officers constituted a special problem. To obtain this group, a large number of qualified women who volunteered for duty were given a Mental-Alertness Test.

Ratings on the qualities of poise, bearing, and leadership of these women were obtained. Those who had dependents or who would have had to leave a child if accepted for service were automatically excluded, as were some specialists who, it was deemed, could render greater service by remaining in their civilian positions. The records of the best two-thirds of the remaining candidates in each corps area were then taken to Washington. At a conference in Washington with psychologists and psychiatrists these records were examined and discussed, and 440 women were selected to be trained for the original cadre of officers of the WAAC. All additional officers were obtained from the ranks of enlisted WAAC personnel who were sent to the officer-candidate school at Des Moines, Iowa, for training.

Civilian applicants for enlistment in the WAAC and in the WAC were tested at once with a classification test constructed especially for the purpose. Women who passed this test and who satisfied other requirements were sent to training centers where additional tests were used to aid in determining their assignment to duty. Among the tests used for this purpose were the Typing Test, the Army General Classification Test, the Mechanical-Aptitude Test, the Clerical-Aptitude Test, the Army Radio Code-Aptitude Test, the General Electrical- and Radio-Information Test, the Driver- and Automotive-Information Test, the Arithmetic Test, and Standardized Oral Trade Questions. As in the case of enlisted men, each member of the WAAC and the WAC was accompanied throughout her Army career by a qualification card summarizing her experience, training, and aptitude-test scores.

SPECIALIZED TRAINING PROGRAM

Selection of men for the Army Specialized Training Program was originally accomplished in conformance with War Department memoranda issued on December 26, 1942, and February 19, 1943. To be eligible for selection, a man had to have obtained a standard score of at least 110 on the Army General Classification Test and to have completed a minimum of nine weeks of basic training. In addition, to be eligible for admission to the basic course of the Specialized Training Program, a man

had to have been graduated from an accredited high school and had to be between the ages of eighteen and twenty-one inclusive. To be eligible for admission to the advanced course of the Specialized Training Program, a man had to have had at least two years of work in a recognized college or university and had to be eighteen years of age, or more. Aviation cadets and officer-candidates were never eligible.

Classification officers at Army installations were authorized to administer an Officer-Candidate Test to eligible men. Those who obtained a standard score of at least 110 on this test and were willing to enter training as privates (regardless of their previous rank) filled out the Personal Data and Interview Form and appeared before a selection board designated to select men for assignment to the Army Specialized Training Program.

Beginning in February 1943, however, these requirements were further modified. For admission to courses in advanced engineering, for example, the qualifying score on the Officer-Candidate Test was raised to 115 and certain course prerequisites were established. One year of college physics and the study of college mathematics through differential calculus were required. The introduction of course prerequisites led to so much difficulty in the qualification of men by the selection boards, the members of which were not ordinarily equipped to evaluate course credits, that on April 9, 1943, a revised procedure for selecting and classifying applicants for specialized training was issued. The qualifying score on the Army General Classification Test was raised to 115 and provision was made for accepting men who had taken the joint Army-Navy College Qualifying Tests, which were first administered under the auspices of the College Entrance Examination Board to 315,962 high school seniors and graduates on April 2, 1943. In addition, Specialized Training and Reassignment (STAR) Units were established for the purpose of receiving, housing, classifying, and instructing personnel chosen by field selection boards as generally qualified for the Army Specialized Training Program.

At first, the classification of a man in a Specialized Training and Reassignment Unit was accomplished on the basis of a personal interview together with data recorded on the Soldier's

Qualification Card and the Personal Data Form plus the man's score on the Officer-Candidate Test. On April 15, 1943, a memorandum from The Adjutant General's Office prescribed the administration of subject-matter examinations and set qualifying scores for admission to various courses in the Specialized Training Program. Later, the qualifying scores were rescinded and local norms were used with the examinations. In general, an effort was made to assign each man to the type of course for which he was best fitted at the highest level for which he could qualify.

The academic year included four twelve-week terms at the end of each of which objective examinations in subject-matter fields were administered to assess the results of instruction. These examinations were constructed by the Personnel Research Section of The Adjutant General's Office. The results of the tests were used as one basis for judging the quality of instruction in the various colleges and universities at which the training units were located. After November 1943 individual test scores were reported to instructors and could form part of the basis for assigning marks.

At the end of each term, the disposition of men in the training courses was considered. Some of them were assigned to the next higher term and others were recommended for an officer-candidate school. It was also possible to transfer men back to the troops. When a course of training had been completed, efforts were made to place the men in the Army where they could make best use of their specialized knowledge and thus make the largest possible contribution to the war effort. Changing needs of the Army and difficulties encountered with quotas often prevented men from being assigned to stations in which they could make adequate use of their specialized training.

SPECIAL GROUPS

In addition to the routine selection of men for ordinary training courses and duty assignments in the Army, many special problems of selection and classification arose. Some of the more important of those handled by the Personnel Research Section of The Adjutant General's Office included the selection of war-

rant officers for more than thirty highly specialized fields, the selection of officer candidates for the large number of officer-candidate schools, the selection of radiotelegraph operators, and the selection of officers for the peacetime Army from the large number of applicants holding temporary commissions in the Army of the United States.

NAVY PROCEDURES

Psychological tests were used in the Navy as early as 1912, but prior to 1923 there were no organized testing programs. In 1924, a Training Division was established in the Bureau of Navigation and a General Classification Test was introduced for use in training stations to select enlisted men for Navy schools. Beginning in 1931, an intelligence test was administered at recruiting stations, and a minimum qualifying score was set for enlistment. Although testing for purposes of selecting and classifying personnel was not emphasized in the Navy in the years prior to 1939, recruits arriving at naval training stations beginning in 1931 were subjected to these examinations:

O'Rourke General Classification Test, Junior Grade (U.S. Navy edition)
Mechanical-Aptitude Test, Junior Grade (U.S. Navy edition)
Standard Test in Arithmetic
Standard Test in English
Standard Test in Spelling
Radio Code Aptitude Test

The scores derived from these tests were entered on the men's service records and could be taken into consideration in regard to their assignments, both initially and on shipboard.

During 1942 so many requests for aid in planning testing programs at various expanding naval installations were received in the Bureau of Navigation (which became the Bureau of Naval Personnel in May 1942) that several psychologists were commissioned for duty in the Bureau and the assistance of the Office of Scientific Research and Development was sought. The first of a series of projects carried out under the auspices of the Office of Scientific Research and Development was initiated immediately thereafter and a Test and Research Section was organized in the Bureau of Naval Personnel.

ENLISTED MEN

Prior to the assignment of men to the Navy from Selective Service, the Navy General Classification Test and the Radio-Technician Selection Test were used to qualify applicants for enlistment. After the assignment of men to the Navy by Selective Service began, seventeen-year-olds who applied at Navy recruiting offices were tested for general aptitude for duty and for special programs for radio technicians, hospital corpsmen, combat aircrewmen, and other duty assignments. Among the quota of men assigned to the Navy by Selective Service was a small percentage of illiterates who had displayed ability to learn. As in the case of men of this type assigned to the Army, special-training units were set up to prepare them for duty status.

Literate men accepted by the Navy were sent to recruit training centers where a series of tests included in the Navy Basic Test Battery was administered to them in order to determine to which naval training school or specialized technical-training course some of them should be sent. Qualifying scores were established for over forty-six training programs for enlisted personnel. The tests included in the Basic Test Battery were:

General Classification Test
Reading Test
Arithmetical-Reasoning Test
Mechanical-Aptitude Test
Mechanical-Knowledge Test (Mechanical Score)
Mechanical-Knowledge Test (Electrical Score)
Clerical-Aptitude Test
Spelling Test
Radio Code Test—Speed of Response

Other tests, such as a Sonar Pitch Memory Test, were administered to some men who had already taken the basic battery in order to qualify them for highly specialized training courses.

Scores derived from all of the tests were recorded on each man's Enlisted Personnel Qualifications Card together with pertinent information regarding his civilian training and experience, his hobbies, and his interests. All of these data were available when the man was interviewed for the purpose of recommending a training or duty assignment. The Enlisted Personnel Qualifi-

cations Card constitutes a part of each individual's service record. Alternative recommendations were also made because of the necessity for meeting quotas based on the personnel requirements of the Navy at the time. About 40 percent of all recruits were sent to elementary naval training schools and about 10 percent were assigned to special-duty assignments, some of them being given ratings or commissions at once. The remaining 50 percent of the men had to be sent directly to ships or shore stations for duty as general-detail hands. An earlier start on systematic classification procedures would have made it possible to reduce greatly the percentage of men assigned to general detail.

By August 1945 over one thousand classification officers and enlisted interviewers had been trained to make proper use of the test scores and other data entered on the Personnel Qualifications Card. These trained men were placed in over one hundred different types of naval installations. In general, if a man to be classified had had sufficient civilian experience in a job for which a counterpart existed in the Navy, he was recommended for assignment to the Navy job for which his civilian experience qualified him. In such cases, the test scores were used only as a general index of the man's ability to learn. For young recruits whose vocational experience was either very limited or nonexistent, aptitude-test scores were of much greater value in making school and duty assignments.

In addition to the selection and classification of recruits for assignment to training schools or to their first duty assignments, classification or reclassification of men was accomplished for many other purposes. At precommissioning centers, for example, the crews for ships in process of construction were balanced to insure a reasonably adequate distribution of talent on each ship. Information about each man assigned to a new ship was supplied to the ship's personnel or executive officer together with technical help in setting up an efficient personnel system aboard the ship. At receiving stations, men going to sea for the first time and experienced men returning to the United States for leave and reassignment were handled. Men recommended by commanding

officers of ships and shore stations for advanced training were screened in the receiving stations. Those found capable of benefiting from advanced training were sent to schools at the level justified by their experience or previous training.

To carry on classification work in the shore establishments in each naval district, personnel officers and enlisted men were assigned to each district. Personnel coming into each district were interviewed to make sure that each man would be assigned to a duty in keeping with his training and experience. District personnel-classification officers were responsible for coordinating classification functions of all activities in their district, including training centers, precommissioning centers, service schools, and receiving stations.

Classification work at service schools centered around counseling students. It was possible to reduce attrition somewhat by reawakening interest; occasionally, transfers to other schools were also arranged. The importance of interest as well as aptitude for success in schoolwork was emphasized. Aboard ship, specially trained classification officers were found only on the largest units of the fleet, but classification service was provided to almost two thousand combat and auxiliary vessels by teams of trained personnel from classification centers.

OFFICERS

To meet the need for nearly 300,000 officers required to man the fleet and the shore stations of the Navy at its peak strength during the war, prospective officers were recruited from many sources. A civilian applicant for direct commissioning was required to fill out an application form which provided information about his background and training. Men obviously lacking the necessary qualifications were weeded out at once; the remainder were given physical examinations and the Officer Qualifications Test and were interviewed by two officers in order to determine the types of assignment in which they would be of most service to the Navy. Men who passed the physical and mental examinations and who seemed to possess skills and abilities needed by the Navy were investigated through references. The applications of the men who were found worthy of final consideration

were forwarded to the Bureau of Naval Personnel where they were reviewed and ultimately either accepted or rejected.

These same procedures were followed in the case of applications from enlisted personnel in the Navy. During the classification process at recruit training centers, enlisted men with the proper qualifications were given the opportunity to apply for commissions. Likewise, at the end of six months' service, any enlisted man could apply for a permanent commission in the Naval Reserve. Temporary appointments to commissioned rank were given to qualified enlisted personnel who were recommended by their commanding officers.

After officers had been selected by means of the procedure described above, the problem of classifying them, training them, and assigning them to appropriate duty stations had to be dealt with. To minimize the amount of training required, every effort was made to place each officer in such a way as to capitalize to the maximum extent on his civilian education and experience. Interviewing officers were assigned to each indoctrination school and each Reserve midshipmen's school to formulate recommendations for the Bureau of Naval Personnel. These interviewing officers acquainted the new officers with the kinds of duty open to them in the Navy and evaluated each officer's qualifications and preferences. All officers took the Officer Classification Test, which included measures of verbal facility, mechanical aptitude, mathematical ability, and ability to visualize spatial relations. Tests to measure aptitude for training in such highly specialized fields as radar or sonar were given to officers who, it seemed, might be candidates for training in those fields. By the time the officer was interviewed personally, a good deal of information concerning him was available to the interviewer. The officer's test scores, his school marks, his preferences regarding an assignment, and data about his background and training were known to the interviewer as he gauged the officer's appearance, physique, and general personality. On the basis of all these considerations and the needs of the Navy at the time, a recommendation for assignment concerning each officer was made to the Bureau of Naval Personnel.

The fundamental data for officer classification were filed in

the Officer Qualifications Record Jacket which followed the man from station to station throughout his Navy career. It formed the basis for the initial classification and subsequent reclassification of each officer. In addition to those in the indoctrination schools and the reserve midshipmen's schools, classification procedures were carried on in operational training centers, aboard ship, and in ports of entry for men returning from sea duty. To facilitate the handling of the data, information on officer qualifications was punched on Hollerith cards.

The procurement of naval aviators, like the procurement of aviation cadets in the Army Air Forces, was accomplished mainly through examining boards in which civilian applicants were examined. In the selection boards for naval aviators, the applicants were checked for the requirements of age, height, weight, and schooling. If the applicant passed a rigorous Flight Physical Examination, he was given psychological tests beginning with the Aviation Classification Test. This test was designed to measure the applicant's ability to learn and to understand directions given in verbal form. Men who obtained a passing score on this test next took a test of mechanical comprehension and filled out a Biographical Inventory. Scores from these two tests were combined to provide a Flight Aptitude Rating. This rating, expressed in a nine-point scale, was designed to predict the applicant's chances of succeeding in flight training. That it does so is indicated by data which show that a much larger proportion of men who attained Flight Aptitude Ratings of E (the lowest category) failed in flight training than of men who attained a rating of A (the highest category). Recent studies of combat performance indicate that the selection of men who are most likely to succeed in flight training does not militate against the selection of men who are most apt to be rated high in combat performance.

WAVE PERSONNEL

The selection and classification procedures employed with WAVE personnel were essentially similar to those used in selecting and classifying male personnel in the Navy. Applications for commissions in the Women's Reserve were accepted only

from civilian women who were between the ages of twenty and forty-nine inclusive and who were either college graduates or had completed two years of college work plus at least two years of successful business or professional experience. Applicants who met these requirements and who possessed skills needed by the Navy were given physical examinations and were interviewed by a selection board. The Officer Qualifications Test used for testing certain groups of male officers was also administered to the applicants. Women who qualified were ordered to preliminary training at the naval reserve midshipmen's school located at Smith College, Northampton, Massachusetts, for an eight-week indoctrination course in Navy procedures.

Early in their course of training at Northampton each officer was interviewed individually and a qualifications card was filled out. Recommendations for suitable assignment for the officers were then formulated on the basis of their prewar experience, their academic training, and their preferences. These recommendations were forwarded to the Bureau of Naval Personnel where assignments to available openings were made.

Enlisted personnel were eligible for assignment to officer training after six months' service. Women who demonstrated outstanding ability were recommended by their commanding officers for commissions. About 12 percent of WAVE enlisted personnel were commissioned in this manner.

Applicants for enlisted status in the Women's Reserve had to be between the ages of twenty and thirty-five inclusive and to have had at least two years of high school or business school education. Physical examinations and the Enlisted Qualifications Test were administered to the applicants. Those who could meet the requirements were sent to the naval training school at Hunter College, New York, for a six- to eight-week basic course. In this school, all women were tested with the Navy Basic Test Battery and with other examinations designed to measure aptitudes for special types of duty. Each woman was interviewed individually and recommendations for assignment were made in accordance with standard Navy procedures. The qualifications card which was filled out for each woman followed her throughout her Navy career and served as the basis for reclassification

if that was deemed advisable. In conformance with existing quotas, some women were sent from the training school to more specialized schools. Others were sent to general-detail duty, but this percentage was ordinarily small.

COLLEGE TRAINING PROGRAM

There were three main sources of men for the Navy College Training Program: first, men in the inactive reserve; second, enlisted men on active duty with the fleet and at shore stations; and third, civilians who had completed or who were about to complete their high school training. Men in the enlisted Reserve and already in college under the V-1 and V-7 programs were transferred to the V-12 program when it was started in 1943, and placed on active duty as apprentice seamen. Men in the V-1 program were required to pass a qualifying examination in the spring of 1943 in order to be eligible for V-12 duty. All men were given the opportunity of withdrawing from the college training program by July 4, 1943, if they preferred not to transfer into the V-12 program.

Aviation cadets procured through Naval Aviation Cadet Selection Boards were also included in the V-12 program and at first spent eight months in college before going on to preflight school. Later, the amount of their college training was increased.

The first selection of civilian high school boys and high school graduates was made in the spring of 1943. On April 2 the Army-Navy College Qualifying Tests were first administered under the auspices of the College Entrance Examination Board in schools and colleges throughout the United States. Of the 315,962 boys who took the examination, 123,206 indicated preference for service in the Navy. From this number, only 16,000 were selected for duty in the V-12 program.

The qualifying examination included four types of items: verbal items, reading-comprehension items, mathematics items, and practical-science items. When the qualifying examination was administered subsequently in November 1943 and in March 1944, its content was somewhat different. The reading-comprehension items were dropped out and the mathematics section was enlarged. This change was made in response to complaints that

the original group of students lacked sufficient preparation in mathematics, especially algebra. By altering the internal weighting of the examination, greater emphasis was placed on the initial selection of men on the basis of their knowledge of mathematics.

The top-scoring men on the qualifying examination who met the requirements of age and marital status were told to report at their own expense to the most convenient naval officer procurement agency. Here, each man was given a physical examination and interviewed by two different individuals. His high school record was also evaluated. Both the results of the interview and the evaluation of the high school record were expressed on a scale from 1 to 10. On the basis of all these data, the quotas of men for assignment to the V-12 program were chosen individually by a selection board consisting of one naval officer, one educator, and one civilian. The quota of men assigned to each state was based on the number of white male high school graduates in the state. Because of this quota system, sectional differences in performance on the Army-Navy College Qualifying Tests were reflected in the group of 16,000 men first ordered into active duty for college training. Men in some northern states, not even considered for final selection, attained scores higher than men in certain other states who were chosen for college training.

The method of identifying enlisted personnel already in the Navy for transfer to the V-12 program originally placed a great deal of responsibility on the commanding officers of fleet units and shore installations. Only enlisted men who were high school graduates, who were between their seventeenth and twenty-third birthdays, and who were unmarried and agreed to remain unmarried until they were commissioned were eligible for transfer to the college training program. Within the limitations of quotas assigned to fleet distributions and naval districts, men whose General Classification Test scores were sufficiently high were recommended for college training by their commanding officers on the basis of general ability and officer-like qualities. After experience had shown the desirability of including specific course preparation in the requirements for transfer to the V-12

program, commanding officers were requested to recommend men who had studied algebra and plane geometry for at least a year each in high school.

At the end of the second term of V-12 training, each man was assigned to a specialized curriculum of some sort. First, the men were divided among three groups on the basis of their choice or their previous enlistment in the Marine Corps, the Coast Guard, or the Navy. Within each of these three groups, the men were sorted according to their first preference among the upper-level curriculums, due regard being given to the physical qualifications and other special prerequisites demanded for admission to certain curriculums. The quota for each curriculum was then filled, beginning with the one having the smallest quota, from the list of men who indicated it as their first preference. On this list, the men were arranged in order according to a weighted composite score. Men who were too low on the list for their first preference to be included in the quota were then considered for their second preference, and so forth. After the assignments had all been made, revisions were sometimes required to resolve practical difficulties in housing and transporting men.

Academic failures in the V-12 program were not greater than might normally have been expected. Between July 1, 1943, and November 1, 1943, *of the men who did fail* the following percentages were accounted for by different subject-matter fields:

	Percentage
Mathematics	28.6
Physics	24.9
History	10.1
English	9.3
Chemistry	9.2
Engineering drawing	6.0

The percentages of the failures attributed to six different causes were as follows:

	Percentage
Low mentality	42.4
Lack of application	32.7
Inadequate preparation	13.8
Lack of officer-like qualities	8.8
Physical illness	1.6
Emotional instability	.7

It seems likely that more stringent selection on the basis of aptitude, particularly on the basis of aptitude for work in science and mathematics, might have cut the rate of failure, especially among men from the fleet and from shore establishments.

MARINE CORPS PROCEDURES

The selection and classification procedures used in the Marine Corps represented a combination of materials and methods derived from the Army and the Navy. Consequently, there is little point in discussing them in detail.

ENLISTED MEN

Prior to the spring of 1943, all men entering the Marine Corps were volunteers. Beginning at that time, however, men were allotted to the Marine Corps from Selective Service. Some of these men were immediately discharged because of physical or psychiatric disabilities that were overlooked or nonexistent when the men were examined in the induction stations. A small percentage of the men were illiterates judged capable of useful service after special training. These men were sent to boot camp at San Diego or Parris Island (where all marines receive their basic training) and given instruction in elementary school subjects along with military training at a slower-than-usual pace.

At the recruit depots at San Diego and Parris Island, the men were given a group of tests. They were interviewed individually and qualification cards were filled out. An adaptation of the Army Soldier's Qualification Card was employed for this purpose and the data recorded were similar to those obtained for men entering the Army. The test scores entered on the card were those derived from the Army General Classification Test, the Army Mechanical-Aptitude Test, and a radio code-aptitude test. The Navy Radio-Technician Selection Test was administered to some men for whom it seemed appropriate, and the resulting scores were also entered on the qualification card.

On the basis of test scores, individual preferences, job history, educational background, and personal characteristics, assignments to specific training courses or duty stations were made.

These assignments were generally made on a tentative basis at the beginning of basic training and were reviewed for final action at the end of basic training. Whenever possible, men were placed in duty assignments that made use of skills acquired in prewar employment. Two factors which were most important in preventing assignment to a duty station in which prewar vocational skills would have been immediately applicable were the lack of civilian counterparts for many marine assignments and the fluctuation of quotas for various duty assignments. Sometimes it was necessary to assign relatively unqualified men to a certain duty station because not enough qualified men were available at the moment in the two recruit depots to meet an immediate demand for men.

After completion of basic training, some men were sent to specialized training schools or to Marine infantry training. The Marine Corps made use of many Army and Navy training schools. Minimum scores on the General Classification Test were required for admission to many of these schools. Some men were assigned to air bases for training in Marine aviation as enlisted men.

OFFICERS

Officers required for the wartime strength of the Marine Corps were obtained mainly by three methods. Specialists were recruited directly from civilian life and assigned directly to the work for which they were commissioned after a brief indoctrination course similar to that of the Navy. Other men commissioned directly from civilian life were sent to training schools of various types. No systematic procedures were employed for selecting officers from civilian life. Classification was accomplished on the basis of interviews with special reference to the prewar occupations of the men.

A second method of obtaining officers was that of commissioning enlisted men of outstanding ability in the field, often in the combat zones. A third method consisted in sending qualified enlisted men to officer-candidate school. A minimum General

Classification Test score of 110 and recommendations by commanding officers were the essential requirements for admission to an officer-candidate school.

Marine aviation cadets were admitted to naval aviation cadet training by means of the same selection procedures required for Navy personnel. Marine enlisted men and nonflying officers were permitted to enter flying training if they could qualify. Men in the later stages of naval aviation training were allowed to elect service in Marine aviation if they preferred to serve with that group.

WOMEN MARINES

The initial group of seven women commissioned to organize and direct the Women's Reserve of the Marine Corps were selected by recommendation and interview. About four hundred additional officers were selected for direct commissions from civilian life by means of the same procedures used to select officers for the WAVES. After that, officer candidates were selected from the ranks of enlisted women in the Marines who had been recommended for officer training. At the officer-candidate school, tests were administered to these women, and their classification was accomplished by means of the same procedures used to classify enlisted men and women.

Applications for service as enlisted women in the Marines were made through naval procurement offices or through Marine Corps recruiting offices. At these offices, application forms were filled out, and the applicants were interviewed in order to discover whether they would be likely to be useful and to provide a basis for personality ratings. The same test used to qualify women for service in the WAVES was employed. Physical examinations were also administered. Women who met the initial qualifications were sent to the recruit depot at Hunter College or, later, at Camp Lejeune, North Carolina. Here the Army General Classification Test, the Army Mechanical-Aptitude Test, the Army Clerical-Aptitude Test, the Navy Radio-Code Test, and typing and shorthand tests were given. As in the case of all services, the women were interviewed and qualification cards

were filled out. On the basis of all available data, especially previous work experience and aptitude-test scores, assignments were made within the limits of existing quotas. About half of the women were sent directly to a duty station and the remainder were sent to a school for specialized training.

COLLEGE TRAINING DETACHMENTS

Enlisted men in the Marine Corps were permitted to enter the Navy V-12 College Training Program if they could meet the same requirements as Navy personnel. A certain quota of the V-12 men who had not served in the armed forces previously was permitted to elect service in the Marine Corps. The selection and classification procedures used with these men were the same as those used with all others in the Navy V-12 program.

PROCEDURES IN THE COAST GUARD

ENLISTED MEN

Prior to the outbreak of World War II, applicants for training as apprentice seamen in the Coast Guard were required to have completed the tenth grade. Those who could meet this educational requirement were interviewed to determine their alertness and literacy. Character references were investigated to rule out men with criminal records and undesirable personal habits. During the war, the educational requirements were dropped to successful completion of the eighth grade, and even this requirement was waived for applicants who seemed otherwise qualified. A qualifying test used by the Navy was administered, and certain minimum scores were set for acceptance of applicants as apprentice seamen and as mates and stewards. The need for specialists was so great that machinists, typists, and others, were accepted and immediately assigned suitable ratings.

At the training stations at Curtis Bay, Maryland, and Alameda, California, classification tests and procedures patterned on those used in the Navy were employed.

OFFICERS

Officers of the Coast Guard are trained at the United States

Coast Guard Academy at New London, Connecticut. Prior to the war, young men between the ages of seventeen and twenty-one were admitted to a four-year course at the Academy, which led to the degree of bachelor of science and to a commission as an ensign. Under the pressure of immediate need for a large number of Coast Guard officers during the war, a special four-month course of intensive training was instituted at the Academy, and the graduates of this course were commissioned as ensigns. Men admitted to this special course as Reserve cadets were drawn from two sources. The first source was unmarried civilian applicants between the ages of twenty and twenty-nine inclusive who were college graduates and who had included at least two semesters of mathematics in their college curriculums. Later on, married or unmarried men between the ages of seventeen and thirty-three inclusive were accepted. The second source of Reserve cadets was enlisted men in the Coast Guard with at least three months' service who were recommended by their commanding officers and who were able to obtain a qualifying score on a test of general mental ability.

It is apparent that the selection of officer candidates from civilian life was accomplished largely in terms of educational requirements. Only in the case of enlisted men recommended for officer training was the requirement of passing a qualifying examination introduced. However, extensive studies have been made at the Coast Guard Academy to determine what types of tests could profitably be used to select men for admission. Data based on four consecutive classes of male Reserve cadets indicate that a combination of aptitude tests was developed that would yield composite scores highly useful in selecting men most likely to succeed in the Academy. The data also show that ratings of the men by psychiatrists and psychologists, when combined with the composite aptitude scores, improved slightly upon the usefulness of the composite aptitude scores. The value of ratings based only on personal interviews cannot be determined because the interviewers used aptitude-test scores and personality-questionnaire results in arriving at their ratings. Among 1,177

men, the percentage of failures in successive deciles based on a combination of aptitude-test scores and ratings was as follows:

Decile	N	Percentage
1 (the highest)	106	10.38
2	116	18.10
3	149	23.49
4	101	23.76
5	128	35.16
6	129	45.74
7	120	51.67
8	108	52.78
9	116	70.69
10	104	82.69

Presumably, on the basis of these and other pertinent data, practical use of the selection procedures that have been developed will be made at the Coast Guard Academy.

SPAR PERSONNEL

A Women's Reserve in the Coast Guard was established on November 23, 1942. Shortly thereafter, arrangements were made with the Navy to share recruiting and training facilities. Applicants for enlisted and officer status were accepted if they had no children under eighteen, could pass the Navy physical examination, and were well recommended. For enlisted status, a woman had to be between the ages of twenty and thirty-six, inclusive and to have completed at least two years of high school education. For officer status, applicants had to be between the ages of twenty and forty-nine inclusive and to have graduated from college, or to have completed two years of college work plus at least two years of successful business or professional experience.

To qualify for acceptance at the recruiting station, each applicant had to obtain a certain minimum score on an objective examination which measured general ability to learn and which emphasized verbal facility. Both enlisted personnel and officers were classified during their recruit training. A variety of aptitude tests was administered, and each woman was interviewed

individually. Test scores and personal data were recorded on a suitable card. On the basis of all available data regarding each woman's aptitudes, work experience, personality characteristics, and interests, assignments to duty stations or to specialized training were made.

PROCEDURES IN THE OFFICE OF STRATEGIC SERVICES

During the war, the Office of Strategic Services had to send overseas a large number of men and women for a wide variety of assignments, most of which put a high premium on tact, resourcefulness, and practical judgment. To reduce to a minimum the chances of sending overseas personnel who might break down under pressure, assessment units were established in the United States to evaluate the total personality of each candidate for employment by the Office of Strategic Services. The largest of these assessment units was located at Fairfax, Virginia.

Inasmuch as the men and women proposed for employment by the Office of Strategic Services were ordinarily chosen because they possessed some specialized skill or talent particularly required by the agency, there was little need to test them for general mental ability or for special aptitudes. What was required was a general assessment of their total personalities with a view to weeding out those who might be expected to break down under pressure. In espionage, counterespionage, or propaganda work in foreign countries it was especially necessary to avoid the use of unstable agents.

About 5,500 candidates for work in the Office of Strategic Services were handled in the assessment units, about 2,400 of them at Fairfax. Here, the candidates were rated on their physical ability and on nine aspects of personality. To obtain these ratings, the candidates were subjected to paper-and-pencil tests, were interviewed by trained psychologists, and were put through a series of situational tests of a type suggested by the work of German and British psychologists. The nine aspects of personality on which each candidate was rated on a six-step scale were as follows:

Motivation for assignment
Energy and initiative

Effective intelligence
Emotional stability
Social relations
Leadership
Security
Observation and reporting
Propaganda skills

To measure some elements of effective intelligence, such as mechanical comprehension, paper-and-pencil tests were employed. For most of the other aspects of personality, ratings based on observation and interviews were made by trained psychologists. Questionnaires and personality inventories were employed as interview aids. Projective techniques were freely used. The most interesting problems were the outdoor procedures, such as the bridge-building and stream-crossing situations set up for candidates. The only method of scoring a candidate on his performance in these situations was that of ratings by trained observers. Usually, three psychologists rated each man independently and then reconciled their differences in conference. Improvisations, patterned after the psychodrama, gave the candidates the chance to work out hypothetical interpersonal problems before an audience of fellow-candidates and staff members.

After all the data had been collected in the three and a half days which each candidate spent at Fairfax, a general assessment was made in a staff conference to determine the man's acceptability. An effort was made to validate the assessments, but this proved to be largely futile because of the impossibility of obtaining satisfactory criterion data. Ratings by superior officers and fellow-workers overseas were found to be unsatisfactory for validation purposes so investigations overseas by trained men were initiated to determine the quality of performance of men and women who had been through the assessment units. As yet, none of the findings are available for release.

The selection of personnel for the overseas operations of the Office of Strategic Services constituted one of the most interesting problems that arose in the selection and classification of personnel for the armed forces. In general, mental ability and

specialized skills were relatively unimportant elements in the selection process because the group proposed for overseas service was already highly restricted on those bases. Methods of measuring the personality traits on which it was desired to base selection are less well developed than methods of measuring skills, abilities, and aptitudes. Hence, recourse was had to clinical procedures involving subjective evaluation of behavior in problem situations. The results were highly satisfying to executives of the central and field agencies of the Office of Strategic Services. As yet, however, no objective determination of the success of the procedures employed has been published.

II. IMPLICATIONS FOR CIVILIAN EDUCATION

IMPLICATIONS for civilian education may be derived from the selection and classification procedures used in the armed forces, but a consideration of them indicates that they are not new; they are, generally speaking, principles that have been advocated for many years by educators and psychologists. This is not surprising since the selection and classification procedures in the armed forces were based on well-established principles and carried out by educators and psychologists drawn from civilian life for that very purpose. Nevertheless, the magnitude of the process of selecting and classifying millions of Americans and the vital importance of utilizing manpower in the nation's defense lend force to the implications that may be derived for civilian education.

In presenting these implications, two important considerations have been kept in mind. *First,* the problems of selecting and classifying men and women in the armed forces are in some respects peculiar. The most important considerations in determining a man's assignment in the armed forces were his abilities, particularly his physical abilities, and military need. To be sure, his preferences were taken into account whenever it was conveniently possible to do so, but freedom of choice on the part of the individual was sacrificed in the interests of speed and practical necessity in building up a striking force capable of destroying the enemy. In civilian education the values of scientific procedures in educational and vocational guidance lie not so much in the increased efficiency with which they permit schools and colleges to utilize staff members and equipment as in the more intangible benefits derived from encouraging individuals with exceptional or specialized talents to study and work in fields that match their abilities and in which they can be happy and make their maximum contribution to society as a whole. Within the framework of a democratic society, scientific procedures for identifying and measuring aptitudes must not be used to classify

students arbitrarily and to direct them into fields of endeavor calculated only to permit them to work efficiently. Instead, these procedures should be used to advise and to guide students to make wise educational and vocational choices in the light of valuable data that scientifically constructed instruments are capable of making available to them. Since students cannot reasonably be expected to interpret such data unaided and to relate them meaningfully to realistic educational and vocational goals, competent advisers or counselors should be and must be provided if aptitude-test data are not to be either misused or ignored.

The fact that the problems of selecting and classifying manpower for military duty were different from those associated with the educational and vocational guidance of civilians indicates that procedures used in the armed forces must under no circumstances be copied blindly. Every effort has been made, therefore, in this part of the report to point out adaptations of Army and Navy procedures that are applicable to civilian education and that have some practical significance.

Second, some implications of the selection and classification procedures used in the armed forces are so well known and accepted that a presentation of them could not rise above the level of banality. A good example of the type of data from which an obvious implication may be derived consists of the evidence that individual differences in mental ability were displayed on an unprecedented scale by the men and women tested in the Army and the Navy. To discuss implications of this sort would not result in a useful contribution.

IMPLICATIONS

1. *Men and women of exceptional and specialized talent can be identified and trained.*

Within the practical limitations inherent in any undertaking of such tremendous scope, American manpower was selected and trained in the armed forces solely in the interests of the national welfare. Whether a man received a certain course of training or a particular duty assignment depended on his abilities

and the needs of the services, though in many cases his preferences were taken into account if it was practicable to do so. In peacetime, the requirements for training citizens in the interests of the national welfare are, by comparison, ill-defined. Freedom of the individual to choose his own activities and to map out a career for himself rather than the compulsion demanded by the exigencies of national defense becomes a paramount consideration. Nevertheless, the selection and classification of men and women in the armed forces primarily on the basis of their merit and of the national welfare suggests that in a society becoming more democratic all the time, some systematic, nation-wide procedure be developed for identifying men and women of exceptional and specialized talent and for providing them with the opportunity of appropriate training regardless of their financial resources or those of the community in which they happen to reside.

The phrase "of exceptional and highly specialized talent" is used deliberately to describe the men and women for whom appropriate training ought to be provided. Too often, it has been assumed that higher education should be available only for those who have displayed outstanding ability in one or more of the traditional academic disciplines. Selection tests have ordinarily stressed literary facility and verbal comprehension and reasoning. As can be shown, however, information gathered in the armed forces demonstrates conclusively that for some essential occupations verbal abilities are not particularly important. Highly specialized talent for designing and operating machinery may be largely unrelated to verbal abilities, yet it may be of greater social importance in the modern world to develop talent of this sort than to develop literary talent. Consequently, in designing instruments for the selection of men and women deserving of advanced education, we should make sure that the needs of modern society are fairly represented. The needs of the armed forces were multitudinous and constantly changing, yet a conscientious and moderately successful effort was made to utilize selection and classification tests and procedures that would identify men and women of all the types of ability required to meet military needs. This implies that

the selection of men and women for advanced education in civilian life could be so managed as to reflect the real needs of society. It is evident that the selection procedures would include the use of tests of many unrelated mental abilities and motor skills, as well as instruments designed to reveal individual interests and traits of personality.

By what agencies educational facilities should be provided to insure appropriate training for outstanding men and women is not a proper subject for discussion in this report. Neither is the means by which financial support would be obtained. These are practical problems for which many solutions have been proposed. In some states free, or essentially free, educational facilities of diverse types are now provided at all levels. In others, state scholarships are provided to supplement those available from privately endowed funds. In the case of the armed-forces training schools, financial support was, of course, provided by the federal government so that men and women from every state and community had access to the same educational facilities. The implication is that some form of national scholarships, such as those proposed by President James B. Conant of Harvard University, be provided.[1]

2. *Effective educational and vocational guidance can be provided for students in schools and colleges.*

A basic principle of the Army and Navy personnel systems was that the abilities and skills of every man should be determined and this information should be used to place him in an assignment where he could make his maximum contribution to the war effort. The ultimate objective of Army classification is succinctly stated in Technical Manual 12–425 as "success in battle through the economical and efficient use of personnel." [2] The more immediate objectives of classification are defined as (1) "to facilitate the placement of individuals in the assignment in which they will be of most value to the service," and (2) "to

[1] Conant, "America Remakes the University," *Atlantic Monthly,* CLXXVII (May 1946), 41–45.

[2] U. S. War Department, *Personnel Classification,* Technical Manual 12–425 (Washington: Government Printing Office, 1944), p. 1.

expedite training by utilizing the abilities, skills, and physical capabilities which individuals bring with them from civil life or acquire during their experience in the Army." [3] That the purposes of classification could not be accomplished by testing and interviewing a man once—at the time of his entrance to the Army—was fully recognized. "Classification is a continuing process during the entire period of an officer's or enlisted man's active service." [4]

In civilian education the essential elements of the armed forces selection and classification procedures should be provided. First, trained counselors should be employed in numbers sufficient to permit them to work effectively with a restricted number of pupils. Second, information regarding each pupil should be made available to the counselors. This information should be kept on file systematically, and some sort of durable cumulative record card should be maintained on which it can be conveniently summarized.

To provide information for the counselors, the services of educational psychologists, test technicians, and medical and psychiatric consultants are required. The educational psychologists and test technicians are best equipped to obtain information regarding each pupil's aptitudes and skills. The medical and psychiatric consultants are able to gather and interpret data regarding the pupil's physical and mental health. It is the counselor's task to know when to refer cases to the various consultants, to coordinate their efforts, and to interpret the information provided in such a way that the pupil will come to understand the pertinent facts about himself and their relationship to his choice of a school or out-of-school career.

3. *Tests of aptitudes required for success in various educational and vocational fields can be made available.*

It is a fundamental principle of effective measurement for purposes of selection and classification that the same set of tests be given the entire group of individuals from which differential selection is to be made. Only in this way can comparable scores

[3] *Ibid.*, p. 1.
[4] *Ibid.*, p. 1.

be satisfactorily obtained with which to compare one individual with another. If the problem of selection is merely one of accepting or rejecting men and women for admission to a highly specific course of training, as was often the case in the armed forces, simple short tests may be employed with satisfactory results. But as the number of human abilities and skills required by the course of training is increased, the length and complexity of the selection tests must be increased to maintain efficiency of selection. When the problem is that of deciding in which one of several courses of training or in which one of many vocations an individual would have the greatest probability of success, the length and comprehensiveness of the differential aptitude tests required to provide the basis for a reliable judgment becomes impressive. It may even become frightening unless one keeps clearly in mind the importance of making reliable judgments in such matters and realizes that several hours spent in taking appropriate aptitude tests may save thousands of hours of misdirected training.

A study made by aviation psychologists in the Office of the Air Surgeon illustrates the degree to which it is possible to select men who will be successful in a course of training. If applicants for aviation-cadet training in the summer of 1943 had been accepted without aptitude tests, it would have been necessary to have started 397 men in pilot preflight school in order to have obtained 100 graduates of advanced pilot training schools. That is, only about one-fourth of unselected applicants for aviation-cadet training (in the summer of 1943) could have been expected to get their pilot's wings. Yet training facilities and personnel would have had to be provided for the unsuccessful three-fourths of the applicants until they were eliminated. On the other hand, of the applicants admitted to pilot preflight school in the summer of 1943 who obtained passing scores on the aptitude tests then in current use (the Aviation Cadet Qualifying Examination and the Aircrew Classification Battery), a much smaller percentage had to be eliminated. In fact, to obtain 100 graduates of advanced pilot training schools, it was

necessary to start only 155 *selected* men in pilot preflight school.[5] When it is recalled that tens of thousands of men were being trained for duty as pilots, it is clear that the saving in training facilities and equipment, instructional staff, and manpower achieved by the use of appropriate psychological tests in only one branch of the Army was tremendous.

The aptitude tests used to select and classify aviation cadets for training in the Army Air Forces included a three-hour qualifying examination and an Aircrew Classification Battery that required about six hours of actual working time. On the basis of these nine hours of examinations, applicants for aviation-cadet training were accepted or rejected and recommendations were made regarding their classification as pilots, bombardiers, or navigators. If about nine hours of testing were required to select and classify men with a reasonable degree of accuracy for training in specific courses in the Army Air Forces, it is obvious that at least that much time would be required to provide the information that is needed about the aptitudes of high school and college students in order to provide them with adequate educational or vocational guidance. Yet we sometimes hear a demand for shortening tests of scholastic aptitude. Often these demands are coupled with pleas for tests of greater diagnostic or differential value. One of the clearest implications of the classification testing done in the Army and Navy is that more time should be spent in testing for purposes of differential selection and classification in order to save enormous amounts of time in training as well as in reclassifying men who fail to succeed in hastily made assignments.

4. *Combinations of highly specialized aptitude tests are more effective for purposes of educational and vocational guidance than tests of general intelligence or general learning ability.*

For most purposes in the armed forces selection and classification programs, tests of general intelligence or general learning ability proved to be less efficient than weighted composite scores obtained from highly specialized tests. An illustration of this

[5] For the data on which these statements are based, see P. H. DuBois, ed., *The Classification Program,* chap. v.

tendency is provided by a set of correlation coefficients between scores on certain tests and final grades at the end of a 112-day airplane mechanics course at Keesler Field, Mississippi. These coefficients, which are shown in Table 1, were obtained in the course of a study made in the Personnel Research Section of The Adjutant General's Office.

TABLE 1

PRODUCT-MOMENT CORRELATION COEFFICIENTS BETWEEN SELECTED TEST SCORES AND FINAL GRADES IN THE AIRPLANE MECHANICS SCHOOL AT KEESLER FIELD

Test	N	M_t	σ_t	M_g	σ_g	r_{tg}
General Technical	400	73.72	19.67	81.56	2.95	.54
Trade-Information	535	73.39	22.17	81.82	2.91	.52
Mechanical-Aptitude	453	106.28	15.03	81.53	2.97	.45
General Classification	584	107.10	12.81	81.64	2.91	.41
Surface-Development	437	39.11	13.88	81.91	2.81	.35
Pulley-Bracket Assembly	504	39.22	16.05	81.63	3.03	.34
U-Bolt	504	28.63	9.85	81.63	3.03	.34
Paper Assembly	535	16.86	3.38	81.74	2.90	.33
Nut-and-Bolt	504	26.31	5.88	81.63	3.03	.30
Arithmetic	536	24.95	11.98	81.74	2.90	.21

The data in Table 1 are typical of those found when the prediction efficiency of a test of general learning ability or general intelligence, such as the Army General Classification Test, is compared with that of each one of several tests believed to measure psychological traits involved in a given criterion. Ordinarily, a number of specialized tests may be found that, in combination, are of greater utility than the general test.

These findings are exactly what one should expect to find on the basis of test theory. So-called "intelligence tests" or tests of general learning ability are not likely to provide as efficient or even as accurate prediction of any stated criterion as a set of carefully selected specialized tests, the scores from which are weighted to yield optimum prediction of the criterion. This does not mean that tests of general intelligence or general learning ability have no usefulness whatsoever in educational or vocational guidance. It does mean that they should be superseded by carefully constructed and properly validated special-purpose tests. The science of educational measurement has advanced

far beyond the point where we should be satisfied with the relative inefficiency of tests of general intelligence or general learning ability. A vivid illustration of the efficiency of prediction that may be attained by concentrated efforts is the correlation coefficient of .66 between the Pilot Stanine used to select men for pilot training in the Army Air Forces and graduation or elimination through advanced pilot training.[6] This may be compared with the correlation coefficient of .31 between the Army General Classification Test and the same criterion in the same sample of essentially unselected applicants for aviation-cadet training.[7] The discrepancy in prediction efficiency is not a reflection on the usefulness of the Army General Classification Test for the purposes for which it was originally constructed; it is simply an illustration of the fact that diligent research can provide specialized selection tests far more useful for predicting a given criterion than a test of general learning ability not especially constructed for a particular purpose.

Specific implications regarding the type of test that would be especially useful for educational guidance are provided from data obtained by the Personnel Research Section of the Adjutant General's Office regarding the proposed Separation Reclassification Battery, by the Test and Research Section of the Bureau of Naval Personnel in connection with the validation of their Basic Test Battery, and by the Office of the Air Surgeon concerning the selection and classification of aviation cadets for the Army Air Forces. *First,* these data suggest that for purposes of educational and vocational guidance, a set of aptitude tests should not yield a profile of comparable scores on the tests themselves, but should yield a set of several composite scores so weighted as to maximize the correlation coefficient between each composite-score variable and the criterion for which it is intended to be predictive. Thus, the same set of aptitude tests may be made to provide information concerning the probability of success of an individual in several different courses of study or vocations. In the Army Air Forces, for example, a set of twenty aptitude tests in the Aircrew Classification Battery was employed

[6] *Ibid.,* Table 5. 7.
[7] *Ibid.,* Table 5. 7.

to obtain several weighted composite scores, which were used to predict performance in various specialties, such as assignments as fighter pilots, bomber pilots, navigators, and bombardiers in the aircrews of the AAF, as well as to estimate a man's general officer quality. These five composite aptitude ratings were expressed in comparable units and provided one basis for recommending that an aviation cadet be trained as a pilot, a bombardier, or a navigator. The scores from each of the twenty tests on which the composite aptitude ratings were made were not reported for consideration in classifying aviation cadets, but they were used extensively for research purposes.

Second, both the Army and the Navy made extensive validation studies to determine the practical usefulness of the selection tests that were either developed experimentally or employed for practical purposes. Wide differences in the value of the tests were revealed. These data, together with the interrelationships of the tests themselves, provided the basis for combining the test scores into composite aptitude ratings. They also led to the formulation of hypotheses regarding additional types of mental skills that should be tested to improve the selection and classification process. Personnel in all of the armed forces who were responsible for the development and use of selection and classification tests agree that the progressive modification of selection and classification instruments on the basis of validation data is the most important single aspect of the techniques employed during World War II to develop those instruments. The most striking illustration of the constant changes made in selection tests on the basis of empirical data is the evolution of the Aviation Cadet Qualifying Examination. In the course of publishing seventeen forms of the examination during the war, nine different combinations of items were employed. A tenth combination had been selected for two additional forms which had been assembled for publication by the time Germany surrendered in May 1945.

Third, the most difficult problem in obtaining validation data is often the selection and quantification of a valid criterion. For studying the validity of tests used in the armed forces a wide variety of criteria were employed. Marks in specific training

courses, graduation or elimination from training schools, ratings by fellow-students and faculty members in training schools, and ratings by superior officers on performance in a certain duty assignment were commonly used in validation studies. Efforts to obtain combat validation data were made in the later stages of the war and considerable ingenuity was exercised to devise means of getting a measure of a man's actual efficiency in combat. The Bureau of Medicine and Surgery of the Navy Department, for example, sent psychologists into the combat zones to obtain from pilots the names of men on whom they would most like to fly wing in combat and whom they would least like to have flying wing on them in combat. Each pilot was told of the confidential nature of his replies and asked to name two men in each category. Specific reasons for his choices were sought and recorded. Scores taken at the time of application for aviation-cadet training were then looked up for men named in the two groups. The resulting data could be analyzed to determine whether tests and other information discriminated between the two groups.

In the Army Air Forces the twenty individual tests in the Aircrew Classification Battery and the composite aptitude ratings obtained from the battery were correlated with various measures of combat proficiency, including number of planes shot down, promotions overseas, number of decorations, and ratings by immediate commanding officers. A tendency for some of the test scores to correlate positively with the combat criteria was evidenced. Sufficient numbers of men were used to permit the conclusion that some of the coefficients obtained were significantly greater than zero.

Few of the criteria used in the studies made in the armed forces were sufficiently reliable to permit the validity coefficients to be high. The important implication for civilian educational practice is that, if ingenuity and care are taken to secure criteria that are realistic and truly important, allowances can be made for the unreliability of the test scores and criterion variables when the data are interpreted. To validate a battery of aptitude tests for clerical work, let us say, one of the criterion variables should undoubtedly be ratings made by the immediate

supervisors of a group of individuals working in business offices. In the same manner, the marks given to students in a course intended to train them for clerical work in a business office should be validated by using the same type of criterion. The

TABLE 2

AVERAGE PRODUCT-MOMENT CORRELATION COEFFICIENTS BETWEEN FIRST-TERM COURSE GRADES AND SCORES ON SIX APTITUDE TESTS IN SAMPLES OF NAVY V-12 STUDENTS *

TEST	FIRST-TERM COURSE GRADES				
	English	History	Physics	Mathematics	Engineering Drawing
	N=763	N=761	N=793	N=751	N=763
Verbal section, College Entrance Examination Board Scholastic Aptitude Test	.52	.49	.33	.14	.05
Verbal Reasoning	.46	.46	.42	.20	.20
Quantitative Reasoning	.26	.36	.46	.39	.39
Mathematics section, College Entrance Examination Board Scholastic Aptitude Test	.28	.39	.52	.36	.36
Spatial Visualization	.05	.07	.25	.59	.60
	N=410	N=408	N=439	N=422	N=409
Mechanical Ingenuity	.12	.14	.33	.50	.50

* Weighted averages were computed by means of Fisher's z transformation. Three groups of beginning freshmen at Yale were tested at the time of admission in July 1943, November 1943, and March 1944. The Mechanical-Ingenuity Test was administered to only a few of the July 1943 entrants, and the data for this group have not been used.
The interval between administration of the aptitude tests and the assignment of course grades was one term, or about four months.

resulting data would show whether performance in the work of the course, as judged by the marks, actually corresponded with performance on the job for which the training was designed to fit the students. The necessity for validating tests and school marks against realistic criteria constitutes an important implication of the Armed Services procedures in selection and classification.

5. *A test of fundamental academic aptitudes can be useful in educational guidance.*

Data pertaining to the use of aptitude tests with Navy V-12 students at Yale University [8] and at several other universities show some interesting trends regarding the prediction of school marks in various subject-matter courses by means of aptitude tests. To make their implications plain to individuals not skilled in the interpretation of statistical data, the coefficients in Table 2 of the article by Crawford and Burnham [9] have been combined and arranged in two separate tables. Table 2 shows the rela-

TABLE 3

AVERAGE PRODUCT-MOMENT CORRELATION COEFFICIENTS BETWEEN SECOND-TERM COURSE GRADES AND SCORES ON SIX APTITUDE TESTS IN SAMPLES OF NAVY V-12 STUDENTS *

TEST	SECOND-TERM COURSE GRADES				
	English	History	Physics	Mathematics	Engineering Drawing
	N=676	N=674	N=696	N=581	N=639
Verbal section, College Entrance Examination Board Scholastic Aptitude Test	.49	.48	.30	.19	.15
Verbal Reasoning	.39	.41	.35	.25	.25
Quantitative Reasoning	.10	.28	.41	.50	.42
Mathematics section, College Entrance Examination Board Scholastic Aptitude Test	.29	.29	.47	.45	.46
Spatial Visualization	.19	.01	.25	.20	.48
	N=375	N=373	N=395	N=297	N=338
Mechanical Ingenuity	.03	.05	.33	.29	.46

* Weighted averages were computed by means of Fisher's z transformation. Three groups of beginning freshmen at Yale were tested at the time of admission in July 1943, November 1943, and March 1944. The Mechanical-Ingenuity Test was administered to only a few of the July 1943 entrants, and the data for this group have not been used.

The interval between administration of the aptitude tests and the assignment of course grades was two terms, or about eight months.

Students were eliminated for deficiency in academic work at the end of the first term and again halfway through the second term. Hence, the correlation coefficients between test scores and school marks at the end of the second term were attenuated owing to restriction of range on the basis of first-term final grades and mid-term grades during the second term. Data required to correct for the effects of this restriction of range are not provided by Crawford and Burnham. If the correction could be made, the coefficients in Table 3 would be appreciably higher, the amount of the increase for a given coefficient being largely a function of its magnitude.

[8] A. B. Crawford and P. S. Burnham, "Educational Aptitude Testing in the Navy V-12 Program at Yale," *Psychological Bulletin,* XLII (1945), 301–9.

[9] *Ibid.,* p. 304.

tionships of six aptitude tests given to V-12 students at the time of their admission to Yale and the marks given to those students at the end of the first term, a period of sixteen weeks. Table 3 shows analogous correlation coefficients, using marks given at the end of the second term. A period of thirty-two weeks, therefore, elapsed between the administration of the aptitude tests and the assignment of second-term marks. Furthermore, some of the students were eliminated for academic deficiency at the end of the first term and during the second term. One should, consequently, expect the coefficients in Table 3 to be generally lower than those in Table 2. The fact that they are not lower than they are and that the pattern of relationships remains about the same as in Table 2 increases one's confidence in the possibilities of predicting academic success or failure by means of combinations of appropriate aptitude tests.

Two general implications for civilian education may be drawn from the data: *First,* it is apparently possible to predict degree of success in conventional subject-matter courses with sufficient accuracy to make the procedures of practical utility in educational guidance. This is by no means a new idea; for many years, experts in education and in measurement have advocated the development and use of such aids in guidance. *Second,* it appears possible to make valid differential predictions of attainment between certain academic subjects. To secure effective differential prediction between two subject-matter fields, it is necessary to have aptitude tests each one of which has a much higher correlation with attainment in one field than with attainment in the other. It is clear from Tables 2 and 3 that the verbal section of the CEEB Scholastic Aptitude Test and the Spatial Visualization Test do correlate quite differently with course grades in English and engineering drawing. To prove that aptitude tests given at the time of entrance to V-12 training at Yale could be used to predict more accurately than chance would permit whether individuals would obtain higher grades in English or in engineering drawing, it would be necessary to prove that the algebraic differences between grades in the two subject-matter fields correlate with algebraic differences between the two

scores used for prediction purposes to an extent significantly greater than zero.

The development of a test of fundamental academic aptitudes can and should be done in a systematic manner. The first step is to define the subject-matter courses for which prediction is to be made. It seems likely that certain courses would naturally group themselves together. All courses in science, for example, might be treated together as one criterion variable. The second step is to identify as many as possible of the mental or motor skills required for learning the body of subject matter and the skills included in each criterion variable. The third step is to devise test exercises to measure each one of these mental or motor skills that appears to be of any importance. It is essential that each mental or motor skill that is to be measured be tested as nearly separately as possible. The degree of intercorrelation among these separate tests is not a crucial matter; it is crucial that each skill be measured separately. Taken by itself, lack of correlation among the tests provides evidence only that a broad range of skills is being measured. The fourth step is to obtain multiple regression weights for predicting each criterion variable from a combination of all the original tests. The fifth step is to eliminate from the battery of original tests any test that has no appreciable part in determining those composite scores that appear to be useful for prediction purposes in educational guidance. The remaining tests may then be assigned approximate weights that are convenient for practical use and that do not appreciably lower the prediction efficiency of the composite scores. The sixth step is to express the composite scores in comparable units, verify their prediction efficiency on new samples, establish norms, and compute their reliability coefficients.

To put in concrete form some suggestions for the types of items that might possibly be useful in a test intended for use in educational guidance, the types are here offered as illustrations under the following headings: "verbal ability," "reasoning abilities," "numerical ability," "perceptual ability," "spatial abilities," and "memory." Composite scores obtained from a combination of these types of items would probably be useful

in predicting school marks in English, foreign languages, social studies, sciences, mathematics, shopwork, mechanical drawing, and clerical procedures. It must be emphasized, however, that the items presented are merely illustrative:

VERBAL ABILITY

Recognition-vocabulary items based on words with a literary flavor should be employed to measure the individual's familiarity with and knowledge of words of a specialized kind. A sample item follows: [10]

thesis

A rule
B drama
C proof
D prediction
*E proposition

REASONING ABILITIES

Data obtained by Dr. L. L. Thurstone and information derived from a factorial study of the judgment and reasoning items used in the Aviation Cadet Qualifying Examination of the Army Air Forces indicate that there are several rather unrelated mental skills involved in reasoning ability. It seems likely that at least three of them would have value in a test of academic aptitude, namely, reasoning in reading, deductive reasoning in verbal terms, and arithmetical reasoning.

A sample reasoning-in-reading item follows:

A good servant to Sir Roger is sure of having it in his choice very soon of being no servant at all. As I before observed, he is so good an husband, and knows so thoroughly that the skill of the purse is the cardinal virtue of this life—I say, he knows so well that thriftiness is the support of generosity—that he can often spare a large fine when a tenement falls, and give that payment to a good servant who has a mind to go into the world, or make a stranger pay the fine to that servant, for his more comfortable maintenance, if he stays in his service.

When a tenement falls, Sir Roger evidently

A has to pay a fine.
B fines all of his servants.
*C is entitled to receive a fine.
D gives the tenement to one of his servants.
E is held responsible for the accident.

[10] The correct answer to this and succeeding sample items is indicated by an asterisk at the left of the letter corresponding to the correct choice.

A sample deductive-reasoning item in verbal terms is as follows:

An automobile salesman has an appointment in a city one hundred miles away to which he must travel by train. If the train on which he must travel is late, he will miss his appointment. If the train is not late, he will miss the train. We do not know whether the train is late.

With this information, we can state positively that

*A he will not be able to keep his appointment.
B he will be able to keep his appointment.
C there is no way of telling whether he will be able to keep his appointment.
D he will have to take a later train.
E he will have to wait for the train.

Following is a sample arithmetical-reasoning item:

A truck goes 10 miles on a gallon of gasoline and 60 miles on a quart of oil. If there were 8 gallons of gasoline in the tank and 1½ gallons of oil in the motor, how far could this truck go?

A 70 miles
*B 80 miles
C 90 miles
D 170 miles
E 440 miles

NUMERICAL ABILITY

Computational facility, as defined by Thurstone's well-known *N* factor, may be measured by simple problems in addition, subtraction, multiplication, and division. Reasoning ability should be conscientiously avoided. A sample item is as follows:

Directions:—Blacken space R if the answer given is correct. Blacken space W if the answer given is incorrect. Work as rapidly as possible without making mistakes.

Add: 219
326
197

742

*R ····
····

W ····
····

PERCEPTUAL ABILITY

A Coding Test employed by the Personnel Research Section of The Adjutant General's Office seems to offer a functional approach to the measurement of the kind of perceptual ability required in some clerical operations. Directions and a sample item follow:

Directions:—Look at the first sample word below *(tree)*. Find this word in the line labeled CODE. Its code number, which is shown beneath it, is 46. Now look at the column of numbers under the word *tree*. The number 46 in this column has been underscored by drawing a heavy line between the dotted lines to show that this is the code number of the word *tree*.

In this test underline as quickly as you can the number in each column which is the code number of the word at the head of the column. Work as quickly as you can without making mistakes.

CODE:		*city*	*hand*	*king*	*tree*	
		52	38	75	46	
tree	hand	city	tree	king	hand	tree
38	38	38	38	38	38	38
....	....	....	....	....	....	....
....	....	....	....	....	....	....
46	46	46	46	46	46	46
■	....	....	....	....	....	....
	....	....	....	....	....	....
52	52	52	52	52	52	52
....	....	....	....	....	....	....
....	....	....	....	....	....	....
75	75	75	75	75	75	75
....	....	....	....	....	....	....
....	....	....	....	....	....	....

SPATIAL ABILITIES

Two types of items which measure different aspects of spatial relations and which might reasonably be expected to contribute to the prediction of academic success are mechanical-movements items and hidden-figures items, both of which were studied intensively in connection with their use in the Aviation Cadet Qualifying Examination.

A sample of the mechanical-movements items follows in Figure 3:

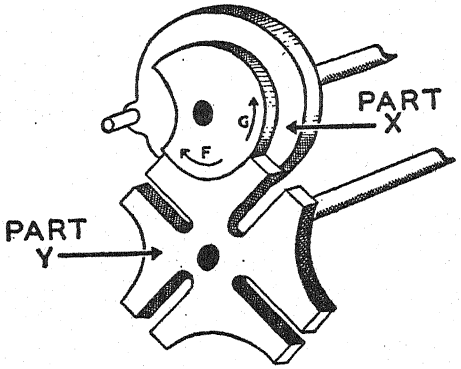


FIG. 3

If part *X* turns at a constant rate of speed, part *Y* turns

A at a constant rate in the same direction as part *X*.
B at a constant rate in the direction opposite to that of part *X*.
C at intervals in the same direction as part *X*, and remains stopped the rest of the time.
*D at intervals in the direction opposite to that of part *X*, and remains stopped the rest of the time.
E first in one direction and then in the other.

Directions for and a sample of the hidden-figures items are as follows:

Directions: To hide objects on the ground from observation by enemy aircraft, the objects are sometimes camouflaged by destroying their natural outlines with the addition of other lines. In each of the exercises that follows, you are to determine which one of five objects, lettered *A, B, C, D,* and *E* at the top of each page, is contained in a camouflaged area. Each camouflaged area is a numbered figure. *One* of the *lettered* objects can be found in each of the *numbered* camouflaged areas. Look at each camouflaged area as you come to it and decide which *one* of the five *lettered* objects is contained in it. The outline of the correct object will always be found right side up in each camouflaged area. Therefore, do not try to rotate the page in order to find it. The outline of the correct object will be the *exact* size and shape of one of the lettered objects shown at the top of the page. In the proper place on your answer sheet, blacken the space corresponding to the *letter* of the object that is contained in each camouflaged area.

Below are the five lettered objects and two sample exercises [see Figure 4].

In sample exercise *00,* you will note that the outline of the object lettered *A* is contained in the lower portion of the camouflaged figure. Therefore,

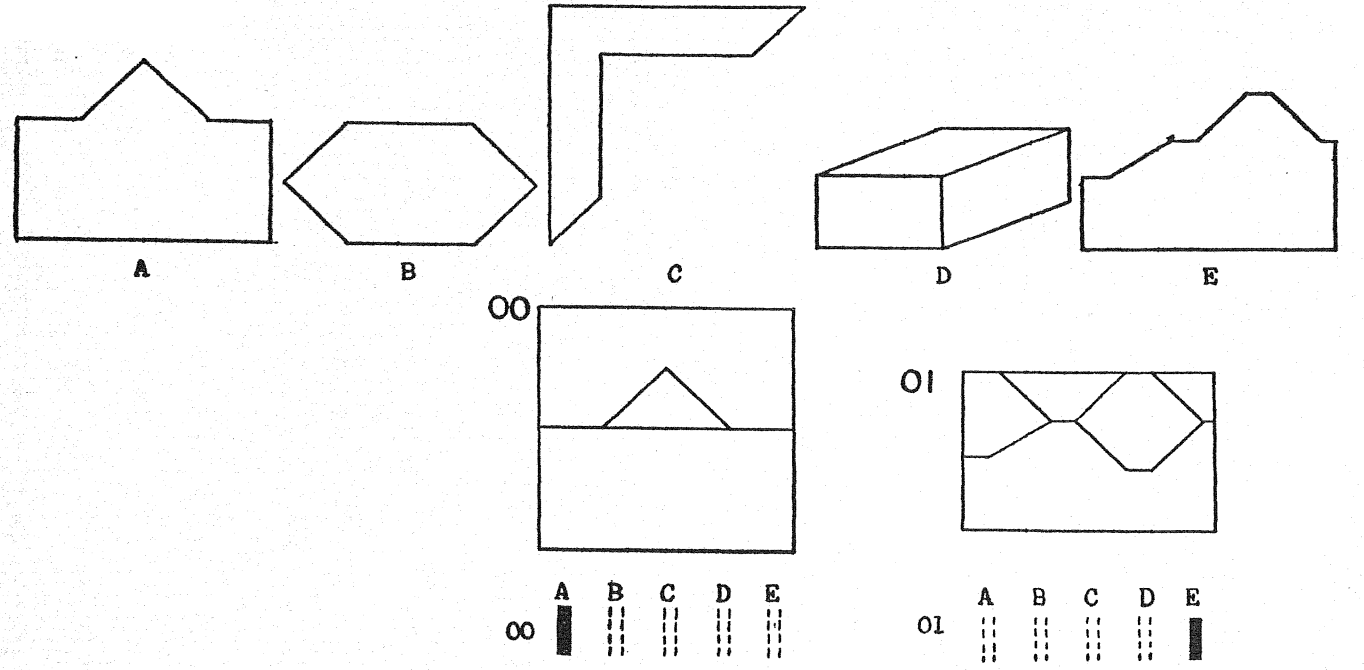


FIG. 4

the answer space below the letter *A* has been blackened in the sample answer spaces.

In sample exercise *01,* you decide which *one* of the five lettered objects shown above is contained in the camouflaged area.

The outline of the object lettered *E* is contained in a portion of the camouflaged area numbered *01.* Therefore, the answer space below the letter *E* has been blackened in the sample answer spaces.

MEMORY

To measure the sort of memory skills required for following directions, a test developed for use in the experimental form of an officer-candidate examination prepared for The Adjutant General's Office may be appropriate. Sample items are not available for reproduction.

It is believed that the types of items listed above under the headings "verbal ability," "reasoning abilities," "numerical ability," "perceptual ability," "spatial abilities," and "memory," together with others suggested by pertinent research, should be administered to junior and senior high school students and validated against school marks and objective measures of achievement in several subject-matter fields. From the experimental battery of tests, a suitable examination for use in educational guidance might be derived. Extensive validation studies should be carried on over a period of several years and critical scores should be obtained for various types of courses in schools and colleges. It is essential that the examination be published by a nationally known organization and that it be made available in convenient form with all required accessory material.

6. *A test of differential aptitudes and interests can be useful in vocational guidance.*

Preparation of a test battery for use in vocational guidance is a more ambitious project than the preparation of a test of fundamental academic aptitudes because the number of mental and motor skills utilized in a wide range of occupations is probably far greater than that utilized in the more common subject-matter fields. The same basic procedures would be employed in constructing both tests, but it is likely that a test of differential aptitudes that would be of practical use in vocational guid-

ance would have to include a wider variety of subtests than a test of academic aptitudes and would, therefore, require considerably more time for administration.

The Personnel Research Section of The Adjutant General's Office had actually assembled and published a Differential Aptitudes Test (Test SG-150a) for use in separation centers when the Army was demobilized. The test included six parts:

Part	Type of Item	Time
I	Word Knowledge	10 minutes
II	Related Figures	20 minutes
III	Arithmetic	20 minutes
IV	Coding	10 minutes
V	Tool Uses	10 minutes
VI	Space Relations	10 minutes
		80 minutes

The choice of tests rested on accumulated data regarding the Primary Mental Abilities Tests of L. L. Thurstone and information regarding a wide variety of tests validated in The Adjutant General's Office during the war. A clear implication of the work accomplished is that a test-building agency of national scope should undertake the systematic development of a battery of tests capable of providing composite scores so weighted as to maximize their value in predicting performance of workers in important groups of occupations.

To supplement information regarding vocational aptitudes, the test battery should include measures of vocational interests. The prototype of useful measures of vocational interests is the Activity-Preference Blank constructed by Truman L. Kelley for The Adjutant General's Office. This blank yields several essentially uncorrelated scores which represent dimensions of vocational interest. The validation of these scores has yet to be made, but the techniques employed in their development are exceedingly promising. It appears unlikely that individuals can make their scores come out the way they want them to as successfully as can be done with most other tests of personality traits and interests. A modified version of the Activity-Preference Blank was administered under the writer's supervision

to a large number of pilots, bombardiers, and navigators in an AAF redistribution station. Experience indicated that revision of the instrument added considerably to its acceptance by the groups to which it was administered. Empirical data are required to determine the extent to which simplification of the blank and of the scoring system would interfere with its usefulness.

Another approach to the measurement of individual interests proved to have considerable merit when it was applied to the problem of measuring the interests of applicants for aviation-cadet training in relation to their success or failure in graduating from advanced pilot training in the Army Air Forces. It consists in determining individual interests by testing many pertinent aspects of general information. A person who knows a good deal about current literature and very little about recent sports events is likely to be a very different sort of person from one who knows a great deal about recent sports events and very little about current literature. Unlike either of them would be the person who knows a great deal about both. In other words, the extent of an individual's knowledge or information about a wide variety of topics may be highly revealing so far as his interests are concerned. Furthermore, the individual is not able to create the false impression that he has a strong interest in some topic because the items are, in a sense, achievement-test items. It is true that he can give the impression that he is less interested in a certain topic than he really is by deliberately marking the wrong answer to an item or by not responding in the case of items to which he actually knows the answer, but this sort of dissembling is unquestionably less common than the kind that accompanies the use of questionnaires or rating scales and is less serious in view of the fact that it can occur only in a negative direction.

Questionnaires and rating scales have the advantage of permitting a direct approach to the measurement of individual interests, thus cutting down the amount of time required to obtain reliable measurement if the individuals tested have no reason to dissemble and do actually answer truthfully. On the other hand, if there is reason to suspect that the individuals

tested might find it to their personal advantage to respond in such a way as to create a false impression of their interests, much less confidence can be placed in the results of questionnaires and rating scales. Unfortunately, the more honest the individual tested, the more he may be penalized in comparison with others for the truthfulness of his responses.

The results of using general-information items to predict graduation or elimination from advanced pilot training in the Army Air Forces were so satisfactory as to imply that this approach to the measurement of vocational interests should be more widely used than it has been. In a large sample of essentially unselected applicants for aviation-cadet training, the biserial correlation coefficient between graduation or elimination from advanced pilot training was higher for scores derived from a general-information test than for scores derived from any other single printed or psychomotor test included in the Aviation Cadet Qualifying Examination or the Aircrew Classification Battery. In fact, the obtained coefficient of .51 was almost as high as the coefficient of .66 for the pilot-aptitude score derived from the entire classification battery. So far as the writer is aware, the efforts of the Cooperative Test Service to develop profiles of interests on the basis of part scores derived from the Cooperative Contemporary Affairs Test have been virtually the only efforts to explore this promising field.

It is important to note that one of the chief values of a single examination designed to provide usable predictive scores for performance in a number of occupations would be its convenience of administration and its ease of interpretation. The test materials available heretofore to guidance counselors have been available only from scattered sources, have been inadequately arranged for mass administration, and have been difficult, if not nearly impossible, to interpret as a coordinated group of measuring instruments. In 1944 Galen Jones referred to this situation when he wrote, "With a few important exceptions, test authors and publishers have been so individualistic, have so completely ignored the needs and convenience . . . of the high school test consumers that many principals have been forced to turn in dismay from tests. . . . The main cause of this confusion and dis-

gust on the part of schoolmen is the lack of comparability in the tests offered for their use. . . . The schools are asked to use a hodgepodge of units and a welter of contradictory 'norms' which are infinitely confusing and discouraging." [11] To correct this situation is the main objective of the suggestions and recommendations in this section of the report. That readily useful differential aptitude and interest scores can be secured from a carefully planned and coordinated battery of tests is indicated by the practical and widespread use of the Aircrew Classification Battery in the Army Air Forces during World War II.

7. *Subjective evaluation of empirical data appears to add little or nothing to the accuracy with which personnel can be selected on the basis of suitable objective tests.*

In both the Army Air Forces and the Coast Guard, subjective evaluations of clinical data were made to supplement the data obtained from batteries of objective tests. At the Coast Guard Academy, the interviewer had the objective test scores of each individual before him as he made his subjective evaluation during the interview. Hence, scores based on the interview were closely related to the weighted composite of the objective test scores. The addition of the interviewer's judgment to the objective test data added very little, however, to the multiple correlation between test scores and final achievement level in the Coast Guard Academy. Similar conclusions were reached by Dunlap and Wantman in a study made for the Civil Aeronautics Authority.[12]

In the Army Air Forces, various subjective evaluations of aviation cadets were made prior to their entrance to preflight school. Interpretations based on the Rorschach Test, observation of performance while taking psychomotor tests, and informal observations during rest periods were some of the bases for making these clinical evaluations. In general, their results were unpromising. No convincing evidence of their practical effectiveness for selecting individuals for pilot training was ever

[11] "Tests and Personnel Work in the High School," *New Directions for Measurement and Guidance* (Washington: American Council on Education, 1944), p. 55.

[12] J. W. Dunlap and M. J. Wantman, *An Investigation of the Interview as a Technique for Selecting Aircraft Pilots* (Washington: Civil Aeronautics Authority, 1944).

obtained. Occasionally, personnel officers or medical officers undertook to make exceptions to the current requirements to pilot training. These exceptions presumably represented individuals who, in the clinical judgment of an officer, could succeed in aircrew training in spite of unsatisfactory performance on the Aircrew Classification Battery, but there is no evidence that men selected in this way were able to succeed in training more often than would have been expected on the basis of their test scores.[13]

Subjective evaluations could not possibly have been made for the vast numbers of men tested, selected, and classified in the armed forces during World War II. Trained clinicians in the required numbers simply were not available. However, it appears that when clinical techniques were applied experimentally to small groups of men, the results of these applications imply that for use in educational and vocational guidance in civilian education they would prove to be ineffective by comparison with the results to be expected from the application of well-rounded combinations of carefully chosen objective tests of aptitudes, interests, and personality.[14] This conclusion should by no means be taken to mean that subjective evaluations of clinical data have no place in guidance. They are unquestionably of value in handling individual cases, particularly cases of an unusual character. But the burden of proving their worth in any practical prediction problem rests clearly with those who advocate their use.

8. *The number of separate mental abilities that can be measured is very large.*

Some years ago many psychologists believed that underlying the skills measured by all kinds of tests there must be a small number of basic mental abilities that function in weighted combinations to constitute the skills that can be measured. This point of view still forms the basis for one systematic approach to measurement for purposes of educational and vocational guidance. The idea is that once these few basic mental abilities

[13] R. L. Thorndike, ed., *Research Problems and Techniques,* AAF Aviation Psychology Program Research Reports, No. 3 (Washington: Government Printing Office, 1946), chap. vi.

[14] J. P. Guilford and J. I. Lacey, eds., *Printed Classification Tests,* AAF Aviation Psychology Program Research Reports, No. 5 (Washington: Government Printing Office, 1946), chap. xxiv.

have been identified and measured directly, any desired skill can be measured by the proper weighted combination of the small number of basic mental abilities of which it must be constituted. Therefore, reliable measures of only a few basic mental abilities need be obtained to permit accurate and efficient prediction of an individual's performance in almost any vocation or any school or college course.

The results of testing hundreds of thousands of men in the armed forces and of analyzing these data suggest to many psychologists that the number of basic mental abilities may often have been underestimated. From factorial analyses of many different matrices of intercorrelations obtained as a result of testing aviation cadets in AAF classification centers, factors that have been mathematically determined have been named as indicated in the following list.[15]

Carefulness
General reasoning I
Integration I
Integration II
Integration III
Judgment
Kinesthetic motor
Length estimation
Mathematical background
Mathematical reasoning
Mechanical experience
Memory I
Memory II
Memory III
Numerical
Perceptual speed
Pilot interest
Planning
Psychomotor coordination
Psychomotor precision
Psychomotor speed
Reasoning II
Reasoning III
Social science background
Spatial relations I
Spatial relations II
Spatial relations III
Verbal
Visualization

There is no objective method of determining whether the names attached to the factors discovered in the analyses are accurate descriptions of the mental abilities represented by the factors. In any case, the fact that so large a number of only moderately correlated factors were identified in tests designed to measure some aspect of the ability to learn to fly an airplane suggests that the number of basic mental abilities may be much larger than was formerly believed. If it turns out that the

[15] *Ibid.*, chap. xxviii.

number of basic mental abilities is fairly sizable, the task of identifying and measuring these abilities may take on the formidable dimensions of the task of constructing separate tests designed to measure every important identifiable skill involved in the vocations and the school courses for which prediction of performance is desired, and determining by multiple regression techniques the individual tests that should be retained in a practical battery.

A good deal of work has already been done to isolate basic mental abilities and basic personality traits and interests. Systematic efforts should now be made to coordinate research in the practical applications of tests of these basic abilities to vocational and educational guidance. As mentioned previously, the results of applying the materials and techniques now available to the selection and classification of men and women in the armed forces were gratifying. The implication is clear that a similar coordination of effort would yield measuring instruments of considerable value in guidance.

9. *Regional evaluation of educational outcomes can be carried out on a wide scale.*

The rapidity with which educational programs had to be set up in the armed forces, often under unfavorable circumstances, and the scarcity of trained personnel to act as instructors led to the necessity for periodic evaluation of the results of these educational programs. This was done in both the Army and the Navy college training programs and in many other training activities, especially in the Navy, by means of carefully constructed objective examinations. The progress of individual students in many basic subjects was determined by means of these tests. In addition, comparisons of the amount of progress made by entire classes were made. Officers in charge of training courses in the Army and the Navy have stated that they consider the periodic use of tests of this kind to have been exceedingly helpful in rendering uniform the course of study among many different institutions and in locating instances where learning was not taking place.

It is apparent that any effort to evaluate teaching efficiency in civilian schools and colleges by means of examinations administered uniformly throughout a given city, county, or state would have to be made with great caution and with limited objectives in view. It seems likely, however, that a reasonable approach to the systematic measurement of minimum essentials should be made on a state-wide basis or, perhaps even better, on a regional basis. Almost everyone can be brought to agree that certain basic skills and factual information should constitute one result of the teaching of a few specific subject-matter courses. In the Army or the Navy, if a man were to operate a special type of gun, it might become a matter of vital importance whether he had mastered the skills involved in its operation. Furthermore, when teams of men were brought together from training schools scattered all over the country, each man had to have been taught to perform the same operation in the same way so that, without retraining, the men could work effectively as a team. In civilian education, such uniformity is not always necessary or, indeed, even desirable; but there are large groups of skills that everyone agrees could very well be standardized. The possession of these skills could be measured objectively on a state-wide basis, and schools and individual classes where these skills were not being taught effectively could identify themselves and take appropriate remedial measures.

At present, a number of state-wide testing programs among school and college students are being carried on. The New York State Regents' Examinations are the best-known examples of tests given to all pupils at certain times in their school careers. The Iowa Every-Pupil Tests are administered widely to school pupils throughout Iowa. Nation-wide testing programs are carried on in schools and colleges by the Cooperative Test Service, the Graduate Record Examination, the College Entrance Examination Board, and other organizations.

10. *Objective tests may serve as an aid in selecting instructors.*

Of considerable interest to school administrators are the efforts made in the armed forces to select instructors by means

of tests. One practical problem confronted in the Army Air Forces was the nature of the assignment to be given to aircrew personnel returned from the theaters of operation prior to the end of the war. To determine which men would be most likely to succeed as instructors, efforts were made to develop tests that would correlate positively with their grades in the central instructors schools and with various criteria of teaching proficiency. The results of these efforts were not entirely uniform and only partly successful. Correlations between the instructor-selection scores and various criteria of teaching proficiency in pilot and navigator training were not significantly positive. The selection of bombardier instructors, on the other hand, seemed to have been more successful. Correlation coefficients between bombardier instructor-selection scores and several criteria are shown in Table 4. The data indicate that the instructor-selection scores

TABLE 4

CORRELATION COEFFICIENTS BETWEEN BOMBARDIER INSTRUCTOR-SELECTION SCORES AND SEVERAL CRITERIA OF TEACHING PROFICIENCY

Criterion	N	r
Over-all instructor rating	441	.54
Supervisors' rating	101	.38
Cadets' rating	90	.61
Officers' efficiency rating	101	.38
Bombardier proficiency test	97	.36
Standard phase check	64	.47

were correlated with the various criteria to a degree that offered prospect of practical utility.

It is reasonable to suppose that the development and validation of the examinations used to select teachers for civilian schools might well be pushed forward. Examinations prepared for the National Committee on Teacher Examinations, under the auspices of the American Council on Education, have already been used extensively throughout the country and some preliminary data regarding their validation have been published. A great deal more should be obtained, and the examinations should be revised in the light of these findings. The experience of test development in the armed forces suggests that constant revision

in the light of the best available validation data should be carried on. Such data are often difficult to obtain but one can hardly expect school officials to place confidence in examinations beyond a try-out period unless data regarding their validity confirm the wisdom of using them.

It is probable that personality measures of the type developed by Kelley in the Activity-Preference Blank and by the staff responsible for the construction of the Aviation Cadet Qualifying Examination should be introduced into teacher-selection examinations. Among the most useful of the separate tests in the bombardier instructor-selection score was an Opinion Questionnaire which provided information regarding personality traits.

APPENDIX A

SOME TECHNICAL PROBLEMS OF MEASUREMENT

SPECIALISTS in measurement believe that experience in handling the problems of selection and classification in the armed forces has emphasized the importance of certain considerations in the construction and the use of tests which have been recognized previously but the importance of which may not have been fully realized. A detailed explanation of these implications is outside the scope of this brief report, but a few of them should at least be mentioned.

1. *The importance of obtaining samples of known characteristics can hardly be overemphasized.*

This implication, as stated, is a commonplace for many research workers, especially those concerned with public-opinion polls. But psychologists engaged in test construction and validation have not always fully appreciated the amount of distortion or lack of consistency in test data that may result from systematic bias or lack of representativeness in successive samples. In twenty-seven different samples ranging in size from 171 to 2,891 aviation students, the raw biserial correlation coefficients between graduation or elimination from elementary pilot training in the Army Air Forces and scores on sets of mechanical-comprehension items prepared for use in the Aviation Cadet Qualifying Examination ranged from .06 to .40 with a median at .23. The value of .06 was obtained in a sample of 309 subjects; the value of .40 was obtained in a sample of 171 subjects. Because of the wide fluctuations in number of subjects from sample to sample and the fact that different items of the same general type were employed in successive samples, no rigorous test can be made to determine whether the range of obtained coefficients is greater than could reasonably be expected on the basis of chance alone. These data are cited merely as illustrative of the wide variations that were obtained from successive samples of aviation students

when the samples were not strictly comparable. They emphasize the desirability of using samples of known comparability.

2. *Research on the effectiveness of aptitude testing and training procedures depends on the measurement of a satisfactory criterion.*

The fundamental purpose of selection and classification procedures used in the armed forces was to assign each man to a duty in which he could contribute most to winning the war. The ultimate criterion for judging the effectiveness of the selection and training of fighter pilots, for example, was their performance in combat flying. A research program for developing tests for selecting fighter pilots ought ideally to use as the criterion for judging the value of individual tests and combinations of them a perfectly reliable measure of proficiency in combat flying. Needless to say, this ideal could not be attained or even closely approached in actual practice. If the complex of activities that makes up combat flying for a fighter pilot could be satisfactorily defined in such a manner as to satisfy competent authorities that the definition was reasonably adequate, it could at best be measured with low reliability.

So great is the importance of having a criterion variable which measures the real objective of a selection program that no effort should be spared to obtain quantitative measurements of as many elements of the real objective—the ultimate criterion —as possible even if these measurements can be made with reliability only slightly greater than zero. The attenuating effects of low reliability in the criterion on correlation coefficients with it can be taken care of by using samples of sufficiently large size. It is far better to use a rather unreliable criterion variable that is closely related to the real objective than to use a highly reliable criterion variable that is only slightly related to the real objective. Considerable ingenuity has been exercised in the armed forces to measure criterion variables of real value for indicating performance in combat. The procedure used by the Bureau of Medicine and Surgery in the Navy Department has already been mentioned.

The importance of defining and measuring a satisfactory criterion variable applies quite as forcefully to the evaluation

of teaching methods and materials as to the evaluation of aptitude and achievement tests. If the objectives of a course in French, for example, were to be stated clearly and measuring instruments devised to measure them, or close approximations of them, in such a way that competent authorities in the field of teaching French were in substantial agreement regarding their adequacy, the task of determining which teaching methods were most efficacious under stated conditions would become straightforward. If the objectives of teaching French are not always the same, which is probably true, the teaching methods best suited to attain any weighted combination of the objectives could be ascertained. The fact that the objectives of language courses given by the armed forces were so clearly and specifically defined goes far to explain why the learning that took place was sufficiently great as to surprise civilian educators.

3. *Validation data based on curtailed distributions may be markedly distorted.*

When a group of men is selected by means of a certain test so that no one below a certain critical score is sent into training, correlation coefficients between scores on the selection test and a criterion variable, such as graduation or elimination from the course, are attenuated to an extent dependent on the proportion of the original group of men who were rejected. That the effects may become very serious are shown by the data in Table 5. When about 87 percent of the applicants for aviation-cadet training were excluded on the basis of the AAF Qualifying Examination and the pilot-aptitude score derived from the Aircrew Classification Battery, the correlation coefficients between certain test scores and the criterion were markedly reduced. Furthermore, because of the differential effect of the selection variables, the validity coefficients were reduced to varying degrees.

Formulas that are entirely appropriate for correcting biserial correlation coefficients for restriction of range under the circumstances that arise in practical work have not been developed. Empirical tests of approximations to the correct procedures that were widely used for research purposes in aviation psychology in the Army Air Forces suggest that these approximations are sufficiently close as to be serviceable. The correction

formulas used most commonly in the Army Air Forces are listed by Thorndike in a report titled *Research Problems and Tech-*

TABLE 5

BISERIAL CORRELATION COEFFICIENTS BETWEEN SCORES ON CERTAIN SELECTION TESTS AND THE CRITERION OF GRADUATION OR ELIMINATION FROM ADVANCED PILOT TRAINING IN THE ARMY AIR FORCES

TEST	CORRELATION COEFFICIENT IN SAMPLE OF	
	Essentially Unselected Applicants for Aviation-Cadet Training	Applicants Admitted to Pilot Training
	(N=1036)	(N=136)
Arithmetic Reasoning	.27	.18
Complex Coordination	.40	−.03
Finger Dexterity	.18	.00
Instrument Comprehension II	.45	.27
General Information	.46	.20
Mechanical Principles	.44	.03

niques.[1] These were based on the fundamental presentation by Pearson in 1903.[2] A new technique for estimating the biserial correlation coefficient in the unselected population from data obtained in a selected sample has been provided by Gillman and Goode.[3] Burt[4] and Brogden[5] have also presented formulas for use in correcting correlation coefficients for restriction of range. Davis has published a brief note on correcting reliability coefficients for range when restriction is accomplished on the basis of a correlated variable,[6] and Kaitz has presented a variant of one of Davis' equations.[7]

[1] Thorndike, ed., *Research Problems and Techniques*, chap. v.

[2] K. Pearson, "Mathematical Contributions to the Theory of Evolution—XI. On the Influence of Natural Selection on the Variability and Correlation of Organs," *Philosophical Transactions of the Royal Society of London*, Series A, CC (1903), 1–66.

[3] L. Gillman and H. H. Goode, "An Estimate of the Correlation Coefficient of a Bivariate Normal Population When X is Truncated and Y is Dichotomized," *Harvard Educational Review*, XVI (1946), 52–55.

[4] C. Burt, "Statistical Problems in the Evaluation of Army Tests," *Psychometrika*, IX (1944), 219–35.

[5] H. E. Brogden, "On the Estimation of the Changes in Correlation and Regression Constants Due to Selection on a Single Given Variable," *Journal of Educational Psychology*, XXXV (1944), 484–92.

[6] F. B. Davis, "A Note on Correcting Reliability Coefficients for Range," *Journal of Educational Psychology*, XXXV (1944), 500–2.

[7] H. B. Kaitz, "Comment on the Correction of Reliability Coefficients for Restriction of Range," *Journal of Educational Psychology*, XXXVI (1945), 510–12.

APPENDIX B

TECHNICAL NOTE ON TEST CONSTRUCTION[1]

UNDERLYING THE procedures suggested in Chapter ii of this report for constructing tests of aptitude for purposes of educational and vocational guidance are certain basic principles of test theory. Among these principles is that of using items with reliability coefficients as high as possible. To determine the logical foundations of this principle and to ascertain its practical effect on test construction is the purpose of this note.

THE EFFECT ON TEST VALIDITY OF VARYING THE RELIABILITY OF EQUIVALENT ITEMS

The true correlation coefficient between two variables, X_a and X_b, which *measure exactly the same mental functions* may be written as follows:

$$r_{\infty\infty} = \frac{r_{ab}}{\sqrt{r_{aA}}\ \sqrt{r_{bB}}} = 1, \qquad (1)$$

where: r_{aA} is the reliability coefficient of variable *a*,
r_{bB} is the reliability coefficient of variable *b*,
r_{ab} is the obtained product-moment coefficient.

If the variables are less than perfectly reliable, their obtained correlation will be:

$$r_{ab} = \sqrt{r_{aA}}\ \sqrt{r_{bB}}. \qquad (2)$$

If the variables are equally but not necessarily perfectly reliable, this may be simplified to:

$$r_{ab} = r_{aA} = r_{bB}. \qquad (3)$$

If all the items in a given test are of equal difficulty and measure with equal but less than perfect reliability only a trait designated as *f*, the correlation of each item with trait *f* may be written as follows:

$$r_{if} = \sqrt{r_{iI}}\ \sqrt{r_{fF}}, \qquad (4)$$

where: r_{iI} is the reliability coefficient of each item,
r_{fF} is the reliability coefficient of the measure of trait *f*.

[1]The writer wishes especially to acknowledge the helpful comments on this note that were made by William G. Mollenkopf.

If the variables are expressed as deviations from their respective means, the correlation of n items of the type described above with trait f may be written:

$$r_{tf} \equiv r_{(x_a + x_b + \ldots + x_n)\,(x_f)} = \frac{\Sigma(x_a + x_b + \ldots + x_n)\,(x_f)}{\sqrt{\Sigma(x_a + x_b + \ldots + x_n)^2}\sqrt{\Sigma x_f^2}}. \tag{5}$$

If all of the test items are of equal difficulty, are equally intercorrelated, and are equally correlated with the criterion, this expression may be simplified to:

$$r_{tf} = \frac{n r_{if}}{\sqrt{n + n(n\text{-}1) r_{ij}}}, \tag{6}$$

where: r_{ij} is the correlation of each item with each other item,
t is the total score on the test,
n is the number of items.

Substituting the values in equations (2) and (4) in equation (6), we have:

$$r_{tf} = \frac{n\sqrt{r_{iI}}\sqrt{r_{fF}}}{\sqrt{n + n(n-1)\sqrt{r_{iI}}\sqrt{r_{jJ}}}}. \tag{7}$$

Simplifying:

$$r_{tf} = \frac{n\sqrt{r_{iI}}\sqrt{r_{fF}}}{\sqrt{n + n(n-1) r_{iI}}}, \tag{8}$$

or:

$$r_{tf} = \sqrt{\frac{n r_{iI}}{1 + (n-1) r_{iI}}}\sqrt{r_{fF}}. \tag{9}$$

This result means that if all the items in a test measure with less than perfect reliability only the criterion trait, are of the same level of difficulty, and are equally reliable, the correlation of the test with the trait is a function of the

1) reliability of the trait,
2) reliability of each item,
3) number of items.

Figure 5 has been prepared to show the effect of altering the average reliability coefficient of the items in three different tests on the validity coefficients of the tests. In Tests A, B, and C, each item measures only the criterion trait with reliability equal

to that of every other item and is of the same difficulty level as every other item. The criterion trait is measured with accuracy corresponding to a reliability coefficient of .81. Test A includes one item, Test B includes ten items, and Test C includes

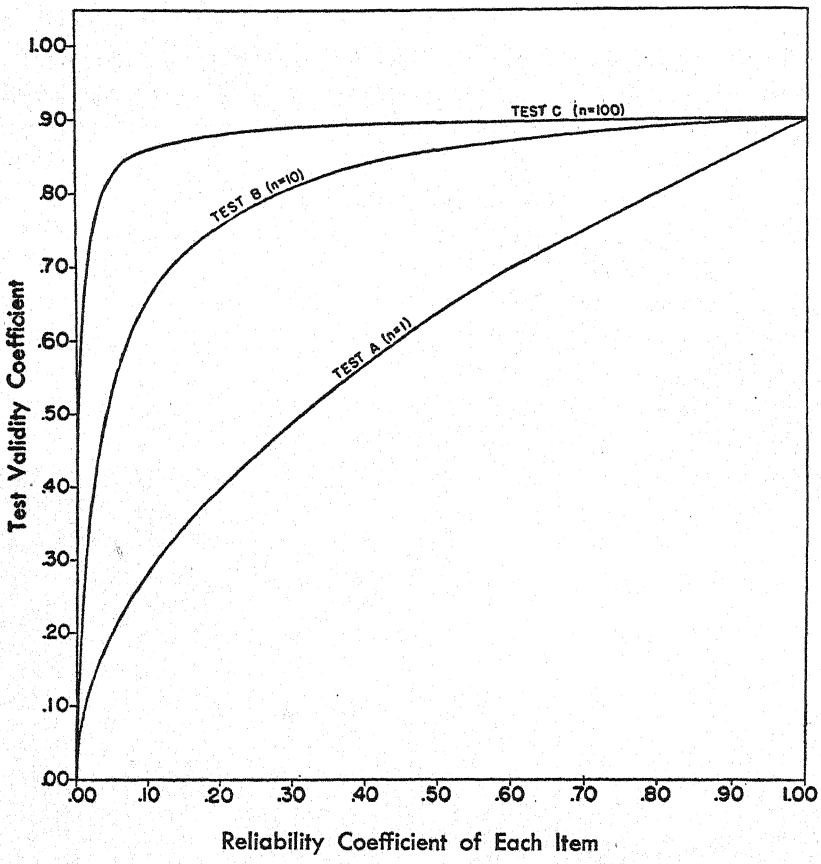


FIG. 5.—Validity coefficient of a test of equally difficult items, each of which measures with less than perfect reliability only the criterion when the latter has a reliability coefficient of .81.

one hundred items. The validity coefficient of each test is shown to increase as the average reliability coefficient of the items in it is increased.

The most striking feature of Figure 5 is the extraordinary rapidity with which the validity coefficient of a test of considerable length, in which all of the items measure only the criterion

variable plus chance, increases with the first few hundredths of average item reliability. The practical implication for test constructors is that when a test composed of homogeneous items is prepared, efforts to maximize average item reliability can pay large dividends in increased test validity. Individual item reliability coefficients as high as .30 are probably not obtainable in practice, so the fact that little gain in the validity of a test composed of ten or more items can be secured with individual items more reliable than that is of little practical consequence. Maximum individual item reliability can be secured by including in a multiple-choice item incorrect choices that are as nearly alike in attractiveness as possible[2] and by editing the item carefully on the basis of expert criticism to insure that it is stated unambiguously and that the choice intended to be the answer is both an adequate answer and incontestably the only correct answer.

The fact that these implications are based on data derived from the use of equally intercorrelated items of equal difficulty and equal reliability probably does not seriously affect their approximate application to the realities of practical test construction. The fact that they are based on items that measure only the criterion variable plus chance might have a greater effect. To explore this possibility, additional data will be considered.

THE EFFECT ON TEST VALIDITY OF VARYING THE RELIABILITY OF ITEMS THAT MEASURE DIFFERENT MENTAL FUNCTIONS

In practice, a criterion variable is likely to be composed of more than one mental function. Hence, to obtain maximum correlation with it, a test must be a weighted composite of measures of several mental functions. Each variable in this composite should, ideally, be a perfectly valid measure of a single mental function included in the criterion. If all of the nonchance variance of the criterion were measured by these variables, the degree of correlation among them would have no effect on the magnitude of the correlation coefficient between any

[2] See A. P. Horst, "The Chance Element in the Multiple-Choice Test Item," *Journal of General Psychology*, VI (1932), 209–11.

given weighted combination of them and the criterion variable. This is also the case whenever any given set of mental functions included in the criterion (but not necessarily comprising 100 percent of its nonchance variance) is measured and is all of the criterion that can, in practice, be measured by any set of test scores.

It is quite true (and not inconsistent with the preceding sentence) that the more nearly uncorrelated are the test scores in any given set, the fewer of them will ordinarily be required to provide any desired level of correlation between a composite of the test scores and the criterion variable. Because of this fact, efforts have sometimes been made to lower the intercorrelations of the separate tests in a set to be used for prediction purposes without regard for the effect of these efforts on the purity of the separate tests. These efforts may be described as misguided because if the intercorrelations of a set of tests are lowered by virtue of making any one or more of the tests measure more than one mental function, the result will almost certainly be to lower the maximum correlation that can be secured between a weighted combination of the tests and the criterion. Under these circumstances, the maximum correlation would *not* be lowered only when the internal weighting of the separate mental functions in a given test just happened to be the same as the weighting those mental functions would be assigned if they were to be measured separately and weighted optimally.

Let us now consider the effect on test validity of varying from zero to unity the reliability of each one of a set of test items that do not all measure the same mental function. If each one of the items in a test is equally difficult, equally intercorrelated with each other item, and measures separately only one of n mental functions included in a criterion variable (trait f) in such a way that each item-criterion correlation coefficient is equal, the correlation of the test with the criterion variable is given by equation (6) when the variables are expressed as deviations from their respective means.

Given certain values for n, r_{ij}, r_{iI}, and r_{fF}, there are limits to the range of the test-criterion correlation coefficient. The lowest value that it can take occurs when r_{ij} is maximized. From

equation (1) it can be shown that:

$$r_{ij} \leqq \sqrt{r_{iI}}\,\sqrt{r_{jJ}}\,. \tag{10}$$

The lower bound of r_{tf} is, therefore, as follows:

$$r_{tf} \geqq \frac{nr_{if}}{\sqrt{n+n\ (n-1)\ \sqrt{r_{iI}}\ \sqrt{r_{jJ}}}}. \tag{11}$$

From equation (1), it can also be shown that:

$$r_{tf} \leqq \sqrt{r_{tT}}\,\sqrt{r_{fF}}\,. \tag{12}$$

This is the upper bound of the test-criterion correlation coefficient. Given fixed values for n, r_{if}, and r_{fF}, equation (6) may be solved for r_{ij} when $r_{tf}=\sqrt{r_{tT}}\,\sqrt{r_{fF}}$. The resulting equation is:

$$r_{ij} \geqq \frac{nr_{if}^2 - r_{tT}\ r_{fF}}{(n-1)\ r_{tT}\ r_{fF}}. \tag{13}$$

To show the effect on the test-criterion correlation coefficient of a ten-item test produced by varying the reliability coefficient of each of its component items from zero to unity and of varying the intercorrelations of the items at each stated level of reliability over the entire range possible, Figure 6 has been prepared. As a starting point, ten items of equal difficulty with reliability coefficients of .16 and validity coefficients of .20 were postulated. The reliability coefficient of the criterion was taken to be .81. By making use of the relationship stated in equation (1), the values were computed for the validity coefficient of each item if its reliability coefficient were varied from zero to unity without altering the mental function tested. Selected values are presented in Table 6.

TABLE 6

RELIABILITY AND VALIDITY DATA REGARDING EACH ITEM POSTULATED

r_{ii} (Reliability Coefficient)	r_{if} (Correlation with Criterion Having Reliability Coefficient of .81)
1.00	.50
.90	.47
.81	.45
.64	.40
.31	.28
.16	.20
.04	.10
.02	.07
.00	.00

Each solid curve in Figure 6 shows the possible range of the validity coefficient of a test of ten items having the characteristics specified. When the items have reliability coefficients of .16, for example, the test validity coefficient can vary from .40 to approximately .57. The validity coefficient cannot drop below .40 because the average item intercorrelation cannot be more than .16. It cannot exceed a value close to .57 becuase that is approximately the point at which the product of the square roots of the test reliability coefficient and the criterion reliability coefficient is at a maximum. The minimum average item intercor-

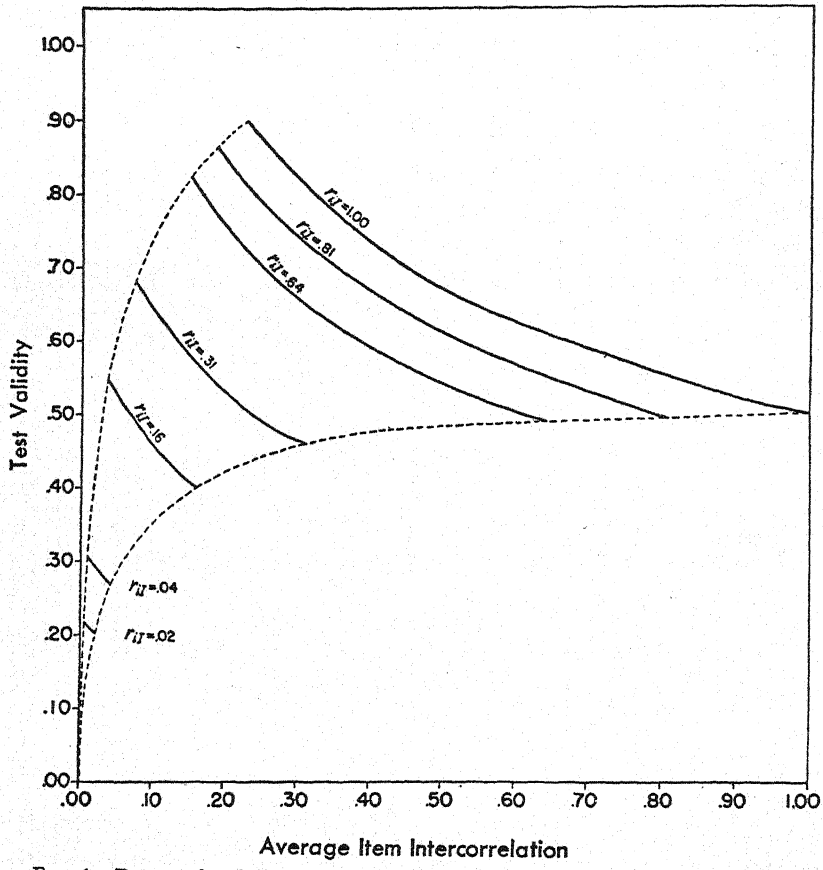


FIG. 6.—Range of validity coefficients that can be obtained from a ten-item test when the reliability coefficient of the criterion is .81 and the individual item reliability is varied from zero to unity.

relation for a test composed of ten items of the type specified would be approximately .037.

There are many ways in which the data presented in Figure 6 can be interpreted. Let us consider some of their implications for test construction:

1. Suppose it was desired to construct a ten-item test having a validity coefficient of not less than .57. If a large number of items of equivalent (and suitable) difficulty and with validity as indicated in Table 6 were available, the goal could be reached by selecting ten of them with reliability coefficients of at least .16. No combination of items with lower reliability could suffice. The more the average item reliability coefficient exceeded .16, the more leeway in the range of average item intercorrelation would become permissible. These observations support the conclusions based on data in Figure 5. The more reliably test items measure any given mental function, the more useful they will be in constructing a test to measure that function.

2. As average item reliability increases, the maximum item intercorrelation that is possible increases. This is shown by the dotted line connecting the bottom of each solid line in Figure 6. The fact that this dotted line rises continuously from left to right demonstrates that the effect of the increase in item intercorrelation on test validity is more than counterbalanced by the effect of the increase in item reliability on test validity. Therefore, it would be futile to expect test validity to increase as a result of a decrease in average item intercorrelation brought about solely by decreasing item reliability.

3. The upper dotted curve running from the lower left-hand corner of Figure 6 to the curve representing the validity of a test composed of perfectly reliable items defines the upper limit of the validity coefficient that is obtainable when the reliability coefficient of each set of ten individual items is varied from zero to unity. Each point on this dotted line represents for a test of ten items of the type specified the magnitude of average item intercorrelation that maximizes the product of the square roots of the test reliability coefficient and the criterion reliability coefficient. Because the highest validity coefficient of a test of ten

items of the type described happens to be obtained when the average intercorrelation of the items is .23, it should not be concluded that this is the optimum value for item intercorrelation except when every one of the ten items is of perfect reliability. With items of lower reliability, then, lower average item intercorrelations are attainable and yield maximum prediction efficiency.

Perhaps the most important implications of the data in Figure 6 may be derived if one thinks of the ten items as ten separate tests, each composed of many items. The best prediction of the criterion is obtained when the average intercorrelation of the tests is necessarily *greater* than zero. Every effort should be made to maximize test reliability by reducing chance factors (not by substituting nonvalid nonchance variance for chance variance). As test reliability is increased, minimum test intercorrelation also increases.

THE EFFECT ON TEST RELIABILITY OF VARYING THE AVERAGE RELIABILITY AND INTERCORRELATION OF INDIVIDUAL ITEMS THAT MEASURE DIFFERENT MENTAL FUNCTIONS

When all of the items in a given test measure with equal reliability the same mental function, the reliability coefficient of the entire test may be obtained by means of the Spearman-Brown formula.[3] However, if the items in a test measure with equal reliability several mental functions, a different formula must be employed for computing the reliability coefficient of the entire test.

Let us say that all of the items in a test are equally difficult, equally intercorrelated with every other item, and that each one measures separately one of n mental functions included in a criterion variable (trait f). Then, if the variables are expressed as deviations from their respective means, the reliability coefficient of the entire test (r_{tt}) may be written as follows:

[3] It is interesting to note that equation (9), which was derived in pages 68–69 includes the Spearman-Brown formula under the first radical sign on the right-hand side of the equation.

$$r_{tT} \equiv r(x_a + x_b + \ldots + x_n)(x_A + x_B + \ldots x_N) = \frac{\Sigma(x_a + x_b + \ldots + x_n)(x_A + x_B + \ldots + x_N)}{\sqrt{\Sigma(x_a + x_b + \ldots + x_n)^2}\sqrt{\Sigma(x_A + x_B + \ldots + x_N)^2}}, \quad (14)$$

where: x_a and x_A, x_b and x_B, $\ldots$, x_n and x_N are pairs of equivalent test items.

Simplifying:

$$r_{tT} = \frac{r_{iI} + (n-1)r_{ij}}{1 + (n-1)r_{ij}}. \quad (15)$$

Equation (15) has been applied to groups of ten equally difficult items that differ in average reliability and average intercorrelation. Each solid curve in Figure 7 represents the possible range of the reliability coefficient of a test of ten items of the specified degree of reliability. From top to bottom, these curves grow progressively shorter because the average item intercorrelation can never exceed the average item reliability coefficient. The effect on the reliability of a test of varying the average intercorrelation of the items in it increases markedly as the average reliability coefficient of its constituent items decreases. The dotted curve in Figure 7 represents the minimum reliability coefficient that can be obtained for a test of ten equally reliable, equally difficult, equally intercorrelated items in which the average intercorrelation takes any value from zero to unity. At the same time, the dotted curve also represents the maximum reliability coefficient that can be obtained for a test of ten equally reliable, equally difficult, equally intercorrelated items of various levels of average reliability. For example, the point on the dotted curve corresponding to an average item intercorrelation of .16 indicates that no test of ten equally reliable, equally difficult items with all item intercorrelations equal to .16 can have a reliability coefficient *lower* than .66. Likewise, no test of ten equally difficult, equally intercorrelated items, each one of which has a reliability coefficient of .16 can have a reliability coefficient *higher* than .66.

If each one of the ten items has a correlation with the criterion variable of .20, the validity coefficient of the entire test will be

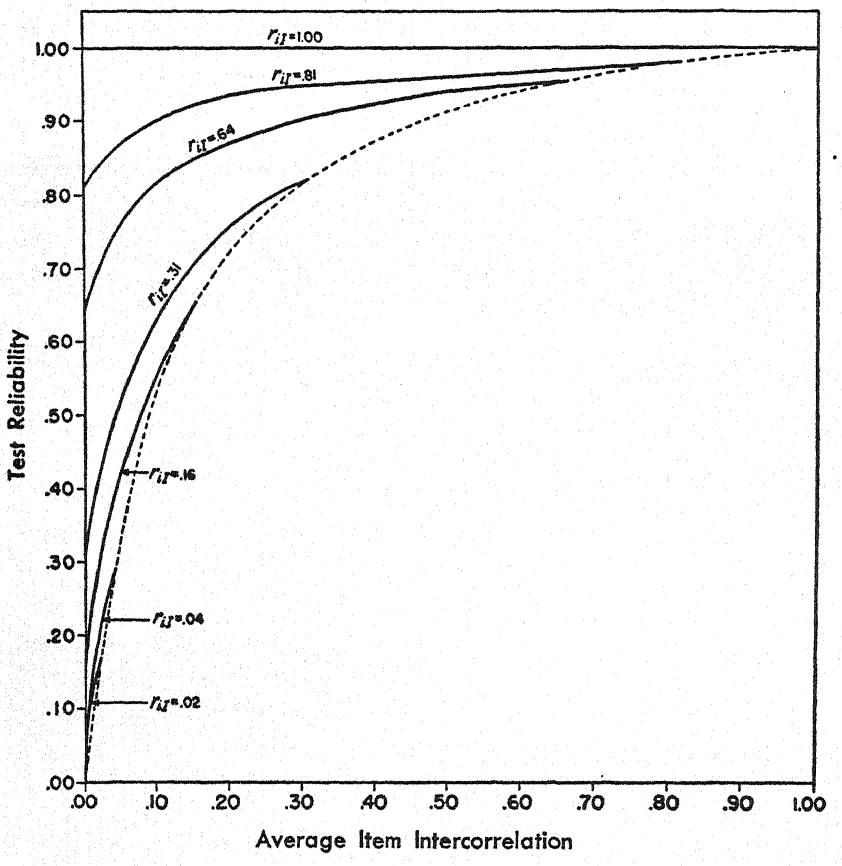


FIG. 7.—Range of reliability coefficients that can be obtained from a ten-item test when the reliability coefficient of each individual item is varied from zero to unity.

.40. The reliability coefficient of this test will, then, be .66.

Now let us suppose that, by careful item construction, it is possible to lower the intercorrelations of all of these items from .16 to .11 without altering either the reliability coefficients or the validity coefficients of the individual items. Then, the validity coefficient of the entire test will rise to .45 and the reliability coefficient will drop to .58. The practical effect of lowering the average item intercorrelation is to increase the *efficiency* of prediction of the test by a small amount and to decrease the *accuracy* of individual measurement. If every item were of 50 percent

difficulty, for example, the standard error of measurement of an obtained score close to the mean would rise from 1.43 to 1.45. This represents a very small decrease in accuracy of measurement and would be more than compensated for by the concomitant increase in test prediction efficiency. Nonetheless, it should be clear that when prediction efficiency is increased solely as a result of decreasing item intercorrelations, accuracy of measurement is sacrificed to some extent. This sacrifice is of no importance if the selection of particular individuals is of no consequence. But when the selection of particular individuals has to be defended in public, the matter may become important. Let us say that an examination is to be used to accept or reject applicants for admission to college or to qualify them for civil service positions. Then it is of great consequence that the rank order of the individuals having scores close to passing be reliable, because it would be embarrassing to explain the use of a test that would not consistently select nearly the same individuals.

Let us suppose that if the energy and care exercised to reduce the average intercorrelation of the items from .16 to .11 had been applied to equalizing the attractiveness of the incorrect choices in the ten items and to removing any trace of ambiguity from them, it would have been possible to increase the reliability coefficient of each item from .16 to .31 without altering the mental function tested by each item. As a result of this change, the validity coefficient of each item would have become .28 and the average intercorrelation would have become .31. Therefore, the validity coefficient of the entire test would have become .46, and its reliability coefficient would have become .82. The increase in item reliability would have caused both the test validity and the test reliability coefficients to rise, thus increasing *accuracy* of individual measurement as well as *efficiency* of prediction.

The practical implication for test construction is obviously that after the types of items found to be most valid for a particular purpose have been identified, every possible effort should be made to construct and edit them so as to purify the mental function measured by each type and to maximize their individual reliability coefficients. Then, the combination of them that yields most efficient prediction can be selected for use.

SELECTED BIBLIOGRAPHY

ALTUS, W. D. "The Adjustment of Army Illiterates," *Psychological Bulletin*, XLII (1945), 461–76.

———. "The Differential Validity and Difficulty of Subtests of the Wechsler Mental Ability Scale," *Psychological Bulletin*, XLII (1945), 238–49.

———. "Racial and Bi-Lingual Group Differences in Predictability and in Mean Aptitude Test Scores in an Army Special Training Center," *Psychological Bulletin*, XLII (1945), 310–20.

ALTUS, W. D., and BELL, H. M. "The Validity of Certain Measures of Maladjustment in an Army Special Training Center," *Psychological Bulletin*, XLII (1945), 98–103.

BELL, H. M., and ALTUS, W. D. "The Work of Psychologists in the Ninth Service Command Special Training Center," *Psychological Bulletin*, XLI (1944), 187–91.

BINGHAM, W. V. "Psychological Services in the United States Army," *Journal of Consulting Psychology*, V (1941), 221–24.

BOBBITT, J. M., and NEWMAN, S. H. "Psychological Activities at the United States Coast Guard Academy," *Psychological Bulletin*, XLI (1944), 568-79.

CARLSON, H. B., and RICH, J. "A Blinker Adaptation of Thurstone's 1943 Code Aptitude Test," *Psychological Bulletin*, XLI (1944), 322–31.

CLEVELAND, A., FAUBION, R. W., and HARRELL, T. W. "Aptitude Tests for Army Weather Observer Students," *Educational and Psychological Measurement*, II (1942), 335–38.

CRAWFORD, A. B., and BURNHAM, P. S. "Educational Aptitude Testing in the Navy V-12 Program at Yale," *Psychological Bulletin*, XLII (1945), 301–9.

DAVIS, F. B. (ed.). *The AAF Qualifying Examination.* Aviation Psychology Program Research Reports, No. 6. Washington: Government Printing Office, 1947.

DEUTSCHER, M. "The Clinical Psychologist in an AAF Mental Hygiene Unit," *Psychological Bulletin*, XLI (1944), 543–47.

DUBOIS, P. H. (ed.). *The Classification Program.* Aviation Psychology Program Research Reports, No. 2. Washington: Government Printing Office, 1947.

EVANS, G. R. "The Army Separation Classification and Vocational Counseling Program," *Occupations*, XXIII (1944), 69–74.

FAUBION, R. W., and BELLOWS, R. M. "Personnel Work in the Army Air Forces: The Classification Division, Army Air Forces Technical Training Command," *Psychological Bulletin*, XXXIX (1942), 643–64.

FELIX, R. H., CAMERON, D. C., BOBBITT, J. M., and NEWMAN, S. H. "An Integrated Medico-Psychological Program at the United States Coast Guard Academy," *American Journal of Psychiatry*, CI (1945), 635–42.

FLANAGAN, J. C. (ed.). *The Aviation Psychology Program in the Army Air Forces.* Aviation Psychology Program Research Reports, No. 1. Washington: Government Printing Office, 1947.

FLANAGAN, J. C. "The Selection and Classification Program for Aviation Cadets," *Journal of Consulting Psychology*, VI (1942), 229–39.

FRANDSEN, A. N., and HADLEY, J. M. "The Prediction of Achievement in a Radio Training School," *Journal of Applied Psychology*, XXVII (1943), 303–10.

FRAZIER, L. G. *The Armed Forces Training Programs of the University of California.* Berkeley, Calif.: The President's Office, University of California, 1945.

GERSTLE, M., JR., WAGNER, R. L., and LODGE, T. "The Inapt Naval Recruit," *U.S. Naval Medical Bulletin*, XLI (1943), 480–91.

GILBERT, G. M. "The Personnel Consultant in an Army Training Center," *Psychological Bulletin*, XLI (1944), 180–86.

GOLDSTEIN, H. "A Malingering Key for Mental Tests," *Psychological Bulletin*, XLII (1945), 104–18.

"A Good Man Is Hard to Find," *Fortune*, XXXIII (March 1946), 92–95, 217–23.

GOODENOUGH, F. L. "The Selection of Candidates for the Officer Candidate School at the Women's Army Auxiliary Corps Training Center," *Psychological Bulletin*, XXXIX (1942), 634–37.

GUILFORD, J. P., and LACEY, J. I. (eds.). *Printed Classification Tests, Parts I and II.* Aviation Psychology Program Research Reports, No. 5. Washington: Government Printing Office, 1947.

HADLEY, J. M. "The Relation of Personal Data to Achievement in a Radio Training School," *Psychological Bulletin*, XLI (1944), 60–63.

HARRELL, T. W. "Aviation Psychology in the Army Air Forces," *Psychological Bulletin*, XLII (1945), 386–89.

———. "Applications of Psychology in the American Army," *Psychological Bulletin*, XLII (1945), 453–60.

HARRELL, T. W., and CHURCHILL, R. D. "The Classification of Military Personnel," *Psychological Bulletin*, XXXVIII (1941), 331–53.

HARRELL, T. W., and FAUBION, R. W. "Selection Tests for Aviation Mechanics," *Journal of Consulting Psychology*, IV (1940), 104–5.

———. "Primary Mental Abilities and Aviation Maintenance Courses," *Educational and Psychological Measurement*, I (1941), 59–66.

HARROWER-ERICKSON, M. R. "A Multiple-Choice Test for Screening Purposes," *Psychosomatic Medicine*, V (1943), 331–41.

HENDERSON, R. W. "Psychological Consultation Service in an Army Specialized Training Program 'STAR' Unit," *Psychological Bulletin*, XLI (1944), 395–99.

HUNT, W. A., WITTSON, C. L., and HARRIS, H. I. "The Screen Test in Military Selection," *Psychological Review*, LI (1944), 37–46.

HUNT, W. A., WITTSON, C. L., HARRIS, H. I., SOLOMON, P., and JACKSON, M. M. "Psychometric Procedures in the Detection of the Neuropsychiatrically Unfit," *U.S. Naval Medical Bulletin*, XLI (1943), 471–80.

HUNT, W. A., WITTSON, C. L., and JACKSON, M. M. "Selection of Naval Personnel with Special Reference to Mental Deficiency," *American Journal of Mental Deficiency*, XLVIII (1944), 245–52.

JARRETT, M. C. *Policies and Procedures Pertaining to the Testing and Classification of Soldiers in the Army Specialized Training Program.* Unpublished Master's Thesis, on file at the George Washington University Library, 1944.

JENKINS, J. G. "Selection and Training of Aircraft Pilots," *Journal of Consulting Psychology*, V (1941), 228–34.

JENSEN, M. B., and ROTTER, J. B. "The Validity of the Multiple-Choice Rorschach Test in Officer Candidate Selection," *Psychological Bulletin*, XLII (1945), 182–85.

JENSEN, M. B., ROTTER, J. B., and HARRISON, R. "Military Psychology in the Armored School," *Psychological Bulletin*, XLII (1945), 91–97.

LAWSHE, C. H., and THORNTON, G. R. "A Test Battery for Identifying Potentially Successful Naval Electrical Trainees," *Journal of Applied Psychology*, XXVII (1943), 399–406.

LAYMAN, J. W. "Reconditioning Program: Army General Hospitals," *Psychological Bulletin*, XLI (1944), 515–18.

LEWINSKI, R. J. "Psychological Services in the Medical Department," *U.S. Naval Medical Bulletin*, XLI (1943), 137–42.

———. "Experiences with the Herring Revision of the Binet-Simon Tests in the Examination of Subnormal Naval Recruits," *American Journal of Mental Deficiency*, XLVIII (1943), 157–61.

LEWINSKI, R. J., and GALWAY, E. J. "Psychological Services at a Naval Retraining Command," *Psychological Bulletin*, XLII (1945), 297–300.

LEWINSKI, R. J., and PENNINGTON, L. A. "Professional Services and Training Requirements of the Psychologist in Class H-V (S) of the Navy," *Psychological Bulletin*, XL (1943), 519–27. See pp. 520–23.

LIMBURG, C. C. "Psychological Work in the United States Maritime Service," *Psychological Bulletin*, XLI (1944), 664–68.

LIPKIN, S. "The Personnel Consultant in an Army Rehabilitation Center," *Psychological Bulletin*, XLI (1944), 524–31.

LOUTTIT, C. M. "Psychological Work in the United States Navy," *Journal of Consulting Psychology*, V (1941), 225–27.

———. "Psychological Examining in the United States Navy: An Historical Summary," *Psychological Bulletin*, XXXIX (1942), 227–39.

MCQUITTY, L. L. "Personnel Selection at an Engineer Replacement Training Center," *Psychological Bulletin*, XL (1943), 509–18.

———. "A Program for the Classification and Training of Retarded Soldiers," *Psychological Bulletin*, XL (1943), 770–79.

MELTON, A. W. (ed.). *Apparatus Tests*. Aviation Psychology Program Research Reports, No. 4. Washington: Government Printing Office, 1947.

MILES, D. W., WILKINS, W. L., LESTER, D. W., and HUTCHENS, W. N. "The Efficiency of a High-Speed Screening Procedure in Detecting the Neuropsychiatrically Unfit at a U.S. Marine Corps Recruit Training Depot," *Journal of Psychology*, XXI, Second Half (1946), 243–68.

MITCHELL, M. B. "The Clinical Psychologist in a Naval Hospital," *Psychological Bulletin*, XLI (1944), 561–64.

MURRAY, H. A., and STEIN, M. "Note on the Selection of Combat Officers," *Psychosomatic Medicine*, V (1943), 386–91.

NEWMAN, S. H., BOBBITT, J. M., and CAMERON, D. C. "The Reliability of the Interview Method in an Officer Candidate Evaluation Program," *American Psychologist*, I (1946), 103–9.

PATTERSON, C. H. "Clinical Psychology in the Army: A Bibliography," *Psychological Bulletin*, XLII (1945), 393–95.

PARTINGTON, J. E., and BRYANT, T. R. "The Personnel Consultant and Psychological Testing at Armed Forces Induction Stations," *American Psychologist*, I (1946), 110–12.

PERSONNEL RESEARCH SECTION, CLASSIFICATION AND REPLACEMENT BRANCH, THE ADJUTANT GENERAL'S OFFICE. "The Army General Classification Test," *Psychological Bulletin*, XLII (1945), 760–68.

———. "The New Army Individual Test of General Mental Ability," *Psychological Bulletin*, XLI (1944), 532–38.

———. "Personnel Research in the Army: I, Background and Organization," *Psychological Bulletin*, XL (1943), 129–35.

———. "Personnel Research in the Army: II, The Classification System and the Place of Testing," *Psychological Bulletin*, XL (1943), 205–11.

———. "Personnel Research in the Army: III, Some Factors Affecting Research in the Army," *Psychological Bulletin*, XL (1943), 271–78.

———. "Personnel Research in the Army: IV, The Selection of Radiotelegraph Operators," *Psychological Bulletin*, XL (1943), 357–71.

———. "Personnel Research in the Army: V, The Army Specialized Training Program," *Psychological Bulletin*, XL (1943), 429–35.

———. "Personnel Research in the Army: VI, The Selection of Truck Drivers," *Psychological Bulletin*, XL (1943), 499–508.

———. "Personnel Research Section, The Adjutant General's Office: Development and Current Status," *Psychological Bulletin*, XLII (1945), 445–52.

———. "Testing as a Part of Military Classification," *Science*, XCVII (1943), 473–78.

Personnel Utilization: Classification and Assignment of Military Personnel in the Army of the United States during World War II. Unpublished report on file in Historical Division, War Department General Staff, Washington, D. C.

PETRULLO, L. "Personnel Procedures in the U.S. Maritime Service," *Psychological Bulletin*, XLII (1945), 177–81.

PSYCHOLOGICAL BRANCH, OFFICE OF THE AIR SURGEON, HEADQUARTERS, ARMY AIR FORCES. "The Aviation Cadet Qualifying Examination of the Army Air Forces," *Psychological Bulletin*, XLI (1944), 385–94.

———. "The Aviation Psychology Program of the Army Air Forces," *Psychological Bulletin*, XL (1943), 759–69.

PSYCHOLOGICAL RESEARCH PROJECT (NAVIGATOR), ELLINGTON FIELD, TEX. "History, Organization, and Research Activities, Psychological Research Project (Navigator), Army Air Forces," *Psychological Bulletin*, XLII (1945), 751–59.

PSYCHOLOGICAL RESEARCH PROJECT (PILOT), RANDOLPH FIELD, TEX. "Psychological Research on Pilot Training in the AAF," *American Psychologist*, I (1946), 7–16.

PSYCHOLOGICAL RESEARCH UNIT NO. 1, NASHVILLE, TENN. "History, Organization, and Procedures, Psychological Research Unit No. 1, Army Air Forces," *Psychological Bulletin*, XLI (1944), 103–14.

PSYCHOLOGICAL RESEARCH UNIT NO. 2, SAN ANTONIO, TEX., AND DEPARTMENT OF PSYCHOLOGY, RESEARCH SECTION, SCHOOL OF AVIATION MEDICINE, RANDOLPH FIELD, TEX. "Research Program on Psychomotor Tests in the Army Air Forces," *Psychological Bulletin*, XLI (1944), 307–21.

PSYCHOLOGICAL RESEARCH UNIT NO. 3, SANTA ANA, CALIF. "Organization and Research Activities, Psychological Research Unit No. 3, Army Air Forces," *Psychological Bulletin*, XLI (1944), 237–45.

PSYCHOLOGICAL SECTION, OFFICE OF THE SURGEON, HEADQUARTERS, ARMY AIR FORCES TRAINING COMMAND, FORT WORTH, TEX. "Psychological Activities in the Training Command, Army Air Forces," *Psychological Bulletin*, XLII (1945), 37–54.

PSYCHOLOGICAL TEST FILM UNIT, SANTA ANA, CALIF. "History, Organization, and Research Activities, Psychological Test Film Unit, Army Air Forces," *Psychological Bulletin*, XLI (1944), 457–68.

RAZRAN, G. H. S., and BROWN, H. C. "Military Psychology: Aviation," *Psychological Bulletin*, XXXVIII (1941), 322–30.

RICHARDS, T. W. "Psychologist in the Naval Hospital," *Psychological Bulletin*, XLI (1944), 565–67.

SEIDENFELD, M. A. "The Special Training Units of the Army," *Psychological Bulletin*, XL (1943), 279–81.

SEPARATION COUNSELING SUBSECTION, CLASSIFICATION AND REPLACEMENT BRANCH, ADJUTANT GENERAL'S OFFICE. "Army Separation Classification and Counseling: I, Objectives and Development," *Psychological Bulletin*, XLII (1945), 379–85.

SHARTLE, C. L. "Occupational and Vocational Counseling of Military and Civilian Personnel During the Period of Post-War Demobilization and the Years Immediately Thereafter," *Psychological Bulletin*, XLI (1944), 697–705.

SOLOMON, P., HARRIS, H. I., WITTSON, C. L., and HUNT, W. A. "Electroencephalography in the Selection of Naval Recruits," *U.S. Naval Medical Bulletin*, XLI (1943), 1310–17.

STALNAKER, J. M. "How the Army-Navy Qualifying Test Was Organized," *School and Society*, LXIII (May 11, 1946), 348–50.

TEST AND RESEARCH SECTION, TRAINING, STANDARDS, AND CURRICULUM DIVISION, BUREAU OF NAVAL PERSONNEL. "Psychological Test Construction in the Bureau of Naval Personnel: Development of the Basic Test Battery for Enlisted Personnel," *Psychological Bulletin*, XLII (1945), 561–71.

———. "Psychological Test Construction and Research in the Bureau of Naval Personnel: Measurement of Achievement in Navy Training," *Psychological Bulletin*, XLII (1945), 769–78.

———. "Psychological Test Construction and Research in the Bureau of Naval Personnel: Validity of the Basic Test Battery, Form 1, for Selection of Ten Types of Elementary Naval Training Schools," *Psychological Bulletin*, XLII (1945), 638–44.

———. "Psychological Testing and Research in the Bureau of Naval Personnel: Work of the Navy's Test and Research Section," *Psychological Bulletin*, XLII (1945), 433–44.

STEVENSON, I. "A Bibliography of Naval Clinical Psychology," *Psychological Bulletin*, XLII (1945), 390–92.

SUPER, D. E. "Clinical Research in the Aviation Psychology Program of the Army Air Forces," *Psychological Bulletin*, XLI (1944), 551–56.

"Testing for Talent," *Fortune*, XXIII (January 1941), 68–70, 95–96.

THORNDIKE, R. L. (ed.). *Research Problems and Techniques*. Aviation Psychology Program Research Reports, No. 3. Washington: Government Printing Office, 1946.

TORPEY, W. G. "A Saga of Sixteen-Year-Old Seamen," *Harvard Educational Review*, XVI (1946), 44–51.

UNITED STATES WAR DEPARTMENT. *Civilian Occupational Classification of Enlisted Personnel.* Technical Manual 12–426. Washington: Government Printing Office, 1944.

———. *Military Occupational Classification of Enlisted Personnel.* Technical Manual 12–427. Washington: Government Printing Office, 1944.

———. *Personnel Classification.* Technical Manual 12–425. Washington: Government Printing Office, 1944.

———. *Personnel Classification Tests.* Technical Manual 12–260. Washington: Government Printing Office, 1942.

"Vocational Guidance for Victory," *Occupations*, XXI (1942), 27–33. Chapter iv.

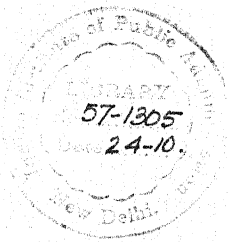
WALLEN, R. "Some Testing Needs in Military Clinical Psychology," *Psychological Bulletin*, XLI (1944), 539–42.

WARNER, G. P. "Each Soldier in the Right Place," *Occupations*, XX (1941), 3–6.

WEITZMAN, E., and BEDELL, R. C. "The Central Examining Board for the Training of Naval Air Cadets," *Psychological Bulletin*, XLI (1944), 57–59.

WITTSON, C. L., and HUNT, W. A. "Psychiatric Selection at the Precommissioning Level," *U.S. Naval Medical Bulletin*, XLV (1945), 621–27.

WITTSON, C. L., HUNT, W. A., and OLDER, H. J. "The Use of the Multiple-Choice Group Rorschach Test in Military Screening," *Journal of Psychology*, XVII, First Half (1944), 91–94.



THE AMERICAN COUNCIL ON EDUCATION

GEORGE F. ZOOK, *President*

A. J. BRUMBAUGH, *Vice President*

The American Council on Education is a *council* of national educational associations; organizations having related interests; approved universities, colleges, and technological schools; state departments of education; city school systems; selected private secondary schools; and selected educational departments of business and industrial companies. It is a center of cooperation and coordination whose influence has been apparent in the shaping of American educational policies as well as in the formulation of American educational practices during the past twenty-eight years. Many leaders in American education and public life serve on the commissions and committees through which the Council operates.

The Commission on Implications of Armed Services Educational Programs began its work in July 1945. It undertakes to identify features of the wartime training and educational programs worthy of adaptation and experimentation in peacetime civilian education of any and all types and levels. It also undertakes to make available to the public well-considered answers to the questions: What should education in America gain from the experience of the vast wartime training efforts? What are the implications for education and the national culture and strength, now and in the future?